Correlation of MicroRNA Expressions with Mutated and Unmutated IgV_H Gene Groups in Chronic Lymphocytic Leukemia

A Thesis

Submitted to the College of Graduate Studies and Research

in Partial Fulfillment of the Requirements

for the Degree of

Master of Science

In the Division of Biomedical Engineering
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5A4
CANADA

By

Yi ZOU

Spring 2005-3-01

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Master of Science degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Chair of the Division of Biomedical Engineering
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5A4
CANADA

ABSTRACT

B-cell chronic lymphocytic leukemia is the most common leukemia in the adult population of Western developed countries. In 2005, an estimated 9,730 adults in the United States will be diagnosed with B-CLL and an estimated 4,600 deaths will occur. B-CLL is a common heterogeneous malignant disease with variable outcome. B-CLL is divided into two groups based on whether somatic hypermutation is observed in the variable region of the immunoglobulin heavy-chain locus (IgV_H). The two distinct groups are named mutated and unmutated. The B-CLL mutated group has a more favorable prognosis than the unmutated group.

Gene expression profiling has been used successfully to decipher the biological and clinical diversity of many leukemias and lymphomas. Recently, other small RNAs (microRNAs) have been shown to be important in hematopoiesis. MicroRNAs are small 20-28 nucleotide RNAs that are believed to control many important cellular and developmental processes by posttranscriptional gene silencing, translational repression, and modulating epigenetic events.

We are interested in whether microRNA expression correlates with the mutational status of IgV_H. This study is significant in the following ways: (1) microRNAs may become surrogate markers for the mutational status of IgV_H of B-CLL, which implies a more rapid diagnostic means as compared to the current practice, and (2) microRNAs,

in the particular context of B-CLL, may play some significant roles in a gene regulatory network that is further responsible for chromosomal abnormalities found in B-CLL.

This thesis presents a study comparing microRNA expression in mutated and unmutated B-CLL groups. Instead of using a genome-wide expression profiling strategy, we selected a specific set of microRNAs based on their chromosome locations and mRNA targets. Specifically, we chose the following eight microRNAs (with their chromosomal abnormalities): **mir16-1 (**deletion 13), **let-7i** (trisomy 12), **mir196-2** (trisomy 12), **mir26a-2** (trisomy 12), **mir-34b** (deletion 11), **mir-125b** (deletion 11), **mir-181C (**trisomy 19), **mir-125a** (trisomy 19). We used solution hybridization assays to monitor the expression of microRNAs. We successfully characterized the microRNA expression in twelve B-CLL patient samples (eight mutated and four unmutated). Among the eight microRNAs examined, three (**mir196-2, mir-125a, mir-125b**) are not expressed in the two B-CLL groups, four (**mir16-1, mir26a-2, let-7i, mir-34b**) have significant differences in expressions over the two groups, and one (**mir-181c**) has no significant difference in expressions over the two groups.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to the following individuals who have made significant technical contributions to my M. Sc. graduate study and thesis: Dr. Ron Geyer who guided me to understand the whole experiment (rationale, design, and procedure) as well as its biochemistry background and implication; Dr. W.F. Dong who guided me to understand the medical background of and implication from the experiment; Dr. Chris Zhang who helped me to understand statistical tools for data analysis relevant to this thesis study and for organizing the thesis in a logic order. The financial assistance to my study from the above individuals through their research grants from various funding agencies (e.g., NSERC, CIHR, etc.) is also greatly appreciated.

I have also received technical advice and helps from Dr. John Decoteau for many interesting discussions on relevant medical problems, and Ying Guo, Susan Bergen, Annette Kerviche, Chris Barreto and Dr. Naoto Takahashi for their excellent instruction in some experimental details. I have found my research place, the Saskatoon Cancer Research Centre, Royal University Hospital, is extremely friendly and exciting owing to many nice individuals (faculty staff, support staff, students) to whom I am grateful but have to omit their names because of too many.

Last, I would like to thank Jean Liu and many others for their friendships and the good times that we shared in Saskatoon, the city I never forget.

Dedicated to my Dad, Mum and Sister

I wish to thank and dedicate this thesis to the members of my family. I cannot complete my studies without their love and support. To my Mom, Caiying Hu, for her never-ending love, to my Dad, Qingyuan Zou, for his confidence and faith in me, for my dear sister, for her encourage and selfless support.

Saskatoon, Saskatchewan

March, 2005 Yi Zou

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BLAST	Basic Local Alignment Search Tool
cDNA	Complementary DNA
CDR	Complementarity-Determining Region
CLL	Chronic Lymphocytic Leukemia
DMSO	Dimethyl Sulphoxide
EDTA	ethylenediaminetetraacetic
FCS	Fetal Calf Serum
FISH	Fluorescence *In Situ* Hybridization
FR	Framework Region
GC	Germinal Centre
GLB	Gel Loading Buffer
IgV_H	Immunoglobulin Gene Variable Region Heavy Chain
miRNA	micro-RNA
mRNA	Messenger RNA
PBS	Phosphate-Buffered Saline
pre-miRNA	Pre-microRNA
pri-miRNA	Primary microRNA
rcf	Relative Centrifugal Force
RISC	RNA Induced Silencing Complex
RNPs	Ribonucleoproteins

RNAi	RNA Interference
rpm	Rotation per Minute
RT-PCR	Reverse Transcription Polymerase Chain Reaction
SAS	Statistical Analysis Software
TBE	Tris Base, Boric Acid, EDTA

CHAPTER 1

INTRODUCTION

1.1 Chronic Lymphocytic Leukemia

Chronic lymphocytic leukemia (CLL) is the most common form of adult leukemia in the Western world (Döhner et al., 2000). CLL is divided into classes, B-CLL and T-CLL, based on their cell of origin (B-lymphoid cell or T-lymphoid cell). Ninety-five percent of CLL cases in North America are B-CLL (Matutes and Catovsky, 1993), and therefore this class of CLL is the focus of this thesis. Patients with B-CLL usually have a gradual onset of symptoms and signs that are related to anemia, infection, and thrombocytopenia (Rozman and Montserrat, 1995). Blood tests showing lymphocytosis are usually the initial sign of B-CLL. Morphology of blood films and flow cytometry are the main methods used for a definitive B-CLL diagnosis. The average onset of B-CLL is 64 years. The outcomes of patients with B-CLL, however, are variable (Cheson et al., 1996). Patients can be grouped based on retrospective outcome studies into two categories; patients with good response to

therapy and long term survival and patients with poor response to therapy and progression to death (Cheson et al., 1996).

At the cellular level, B-CLL cells are monoclonal B-cells that express mature B-cell markers (CD23 and CD5) and that have weak expression of surface immunoglobulins. B-CLL cells have a prolonged life span compared to normal lymphocytes (Lagneaux et al., 1998). B-CLL cells are more resistant to apoptosis, but they do not differ significantly in their rate of cell cycling compared to normal lymphocytes. The gene expression pattern of B-CLL cells suggests that they are related to memory or naïve B-lymphocytes (Küppers et al., 1999).

The clinical course of B-CLL varies considerably. Approximately one third of patients never require treatment and die from causes unrelated to B-CLL. Another one third of patients go through an initial indolent phase that is followed by progression of the disease. The remaining one third of patients have an aggressive disease at the onset and need immediate treatment (Schroeder et al., 1994).

Biological markers for individual prognosis and prospective studies are important for patient treatment and management. Initial efforts to sub-classify B-CLL into distinct groups are based on the following criteria: (i) whether the variable region of the immunoglobulin heavy-chain locus (IgV_H) has undergone somatic mutation, and (ii) whether there is expression of CD38 on the cell surface. Several studies have shown that a more favorable prognosis is associated with B-CLL cells that

have undergone somatic mutation and have low CD38 expression (Damle et al., 1999). B-CLL is commonly divided into two groups based on whether the B-CLL cells have undergone somatic mutation (**Figure 1.1**). Somatic mutation is quantified by measuring the DNA sequence differences between an expressed Ig heavy chain variable (V) region gene (IgV_H) and its corresponding germline DNA sequence. When differences between the IgV_H and germline DNA are greater than 2 %, then the IgV_H is defined as having undergone somatic mutation (Fais et al., 1998).

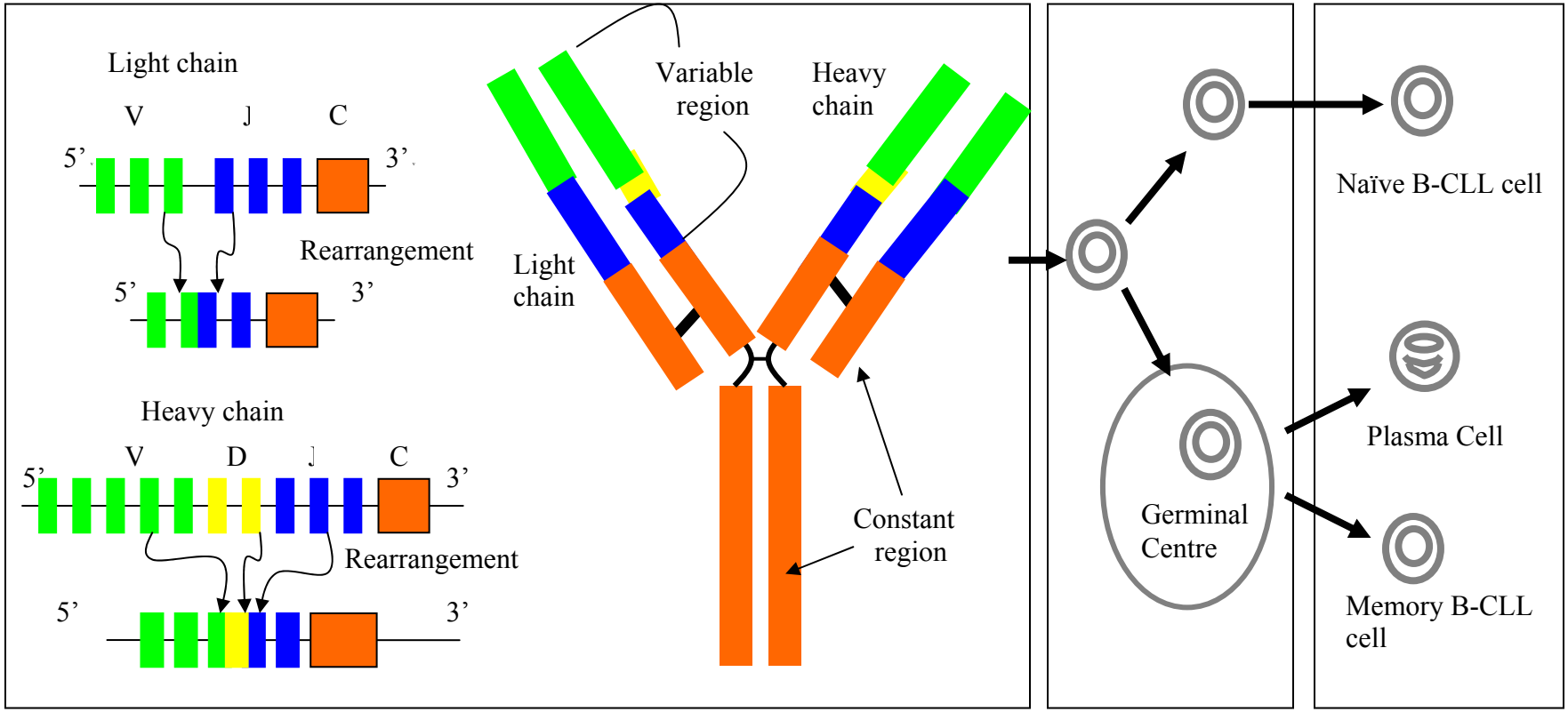

Figure 1.1. Pathogenesis pathway of B-CLL cells. B-cells develop and mature in bone marrow. The immunoglobulin (Ig) heavy chain variable regions of B-cells are encoded in three gene segments: V_H, D_H and J_H. The immunoglobulin light chain variable regions of B-cells are encoded in two gene segments: V_L and J_L. Heavy chain and light chain genes undergo rearrangements during B-cells development. Mature B-cells are referred to as naïve B-cells. If naïve B-CLL cells go through germinal centre, they undergo somatic hypermutation and function as memory lymphocyte, mutated B-CLL cells; otherwise they are unmutated naïve B-CLL cells.

1.2 Prognosis and Biological Cause of B-CLL: A Brief Survey

The development of two different staging systems by Rai et al. (1975) and Binet et al. (1981) made it possible to divide patients with B-CLL into three prognostic groups: good, intermediate, and poor. This division is based on the following signs: lymphocytosis, splenomegaly, lymphadenopathy anemia, thrombocytopenia, and organomegaly. These staging systems do not correspond well with each other. Neither staging system enables physicians to accurately predict which patients in the good-prognosis group will eventually have progressive disease.

Serum levels of β-2-microglobulin, lactate dehydrogenase (Han et al., 1989), soluble CD23 (B-cell membrane protein), and thymidine kinase (Hallek et al., 1999) can help predict disease prognosis. Recent studies found that cytogenetic abnormalities, such as deletions in chromosome 13q14, 11q6 (Stilgenbauer et al., 2000), or somatic mutations in their immunoglobulin heavy-chain genes (Damle et al., 1999), are better predictors of rapid progression and survival. Mutational status of a patient's IgV_H gene is not a routine clinical test due to the cost and complexity. Therefore, many current studies are focused on finding surrogate markers that are independent of or correlated with IgV_H mutational status and can be used to predict disease progression, survival, and treatment. **Table 1.1** lists current markers that are used for B-CLL prognosis, including the Rai and Binet systems and serum level of β-2-microglobulin. Other markers in listed in **Table 1.1** are discussed below.

Table 1.1. Current Prognostic Markers

Staging System or Marker	Clinical Practice	Clinical Trail
Staging(Rai or Binet)	Yes	Yes
β -2-microglobulin	Yes	Yes
Cytogenetics by FISH	Yes	Yes
CD38 status	Yes	Yes
IgV_H mutational status	No	Yes
ZAP-70	Not available	Yes
Thymidine Kinase	No	Yes

Expression of CD38 and ZAP-70 correlates with the presence of unmutated IgV_H genes and an unfavorable clinical outcome (Damle et al., (1999) and Durig et al., (2003)). However, the expression of CD38 and ZAP-70 in both mutated and unmutated B-CLL cells indicates lack of efficacy as a surrogate to classify the mutated and unmutated B-CLL subgroups (Thunberg et al., 2001; Brian et al., 2003)

Genomic aberrations in B-CLL cells are important independent predictors for B-CLL progression and survival. These findings have implications for designing risk-adapted treatment strategies (Döhner et al., 2000). A recent cytogenetic study on B-CLL analyzed survival in relation to chromosomal abnormalities detected using fluorescence *in situ* hybridization (FISH) (Gozzetti et al., 2004). Importantly, this study classified B-CLL into five groups based on the following karyotypes; 17p deletion, 11q deletion, 12q trisomy, normal karyotype, and 13q deletion. The median survival time was correlated with each karyotype. B-CLL disease duration correlated 17p and 13q deletions. 17p showed the shortest survival time (32 months). 13q deletion had longest survival time (133 months). The genomic regions, affected by chromosomal deletions,

amplifications, and less frequently by translocations, still contain unknown tumor suppressor genes and oncogenes.

IgV_H mutation status and the pattern of genomic aberrations have a high predictive value for disease progression and survival in B-CLL patients. However, there are discrepancies in the literature with respect to correlations between genomic aberrations and mutational status of IgV_H genes. For example, B-CLL patients with 13q deletion have a long survival time, however, Kröber et al., (2000) reported that almost half of unmutated B-CLL cases exhibited 13q deletion. This lack of correlation between 13q and mutated B-CLL patients suggests that inactivation of an unknown tumor suppressor in 13q may be important in both unmutated and mutated B-CLL cases. However, correlations between B-CLL mutation status and genomic aberrations are reported by Carsten et al., (2003). In these studies, the favorable prognostic abnormality (13q deletion) is significantly more frequent in the mutated B-CLL group; whereas the prognostic unfavorable abnormalities (17p and 11q deletions) are almost exclusively found in the unmutated B-CLL group. Further, within the unmutated B-CLL group, patients with high-risk genomic aberrations, as defined by 17p and/or 11q deletion, have a significantly inferior survival probability (Ivan and Ian, 2003). Thus, the IgV_H mutational status and genomic aberration likely give complementary prognostic information.

Recently, genomic approaches have been used to classify B-CLL. In 2001, Rosenwald et al,. used lymphochip cDNA microarrays containing 17,856 human cDNAs to profile gene expression in B-CLL samples. They found that the expression of hundreds of other

genes correlated with the IgV_H mutational status in B-CLL, providing insights into the biological mechanisms that lead to the divergent clinical behaviors of B-CLL patients (Rosenwald et al., 2001).

More recently, Calin et al., (2004) used microRNAs (miRNA) microarrays to profile the miRNA expression in B-CLL patients. They found low level expression of mir-16-1 in B-CLL patients harboring deletions at 13q14. They also found a distinct B-CLL miRNA signature composed of five differentially expressed miRNAs (mir-186, mir-132, mir-16-1, mir-102, and mir-29C) that distinguished mutated and unmutated B-CLL groups.

1.3 microRNAs

miRNAs are endogenous 21-24 nucleotide (nt) RNAs that mediate post-transcriptional gene regulation by pairing with the 3' untranslated region of messenger RNAs (mRNA) and repress translation (Moss, 2003). miRNAs are found in mammals, fish, worms, flies, yeast and viruses (Mourelates et al., 2002; Ambros et al., 2003; Aravin et al., 2003; Dostie et al., 2003; Kim et al., 2003). Recent reports suggest a role for miRNAs in development, cell differentiation, apoptosis, and cancer (Baehrecke, 2003; Bartel, 2004; Calin et al., 2004).

miRNA expression in animals involves at least two processing steps (Lee et al., 2002). miRNA precursors are expressed as long primary transcripts (60-110 nt), called the primary microRNA (pri-miRNA), which may be polycistronic. pri-miRNAs are

processed in the nucleus (Lee et al., 2002) by the nuclease Drosha, which yields one or more hairpin precursor sequences (pre-miRNAs or stRNA, small temporal RNA). Pre-miRNAs are exported to the cytoplasm by means of the Exportin-5 pathway (Lund et al., 2004). Cytoplamic pre-miRNAs are processed by the nuclease Dicer. One strand from the RNA duplex is complexed with a protein complex containing argonaute 2, Gemin3, and Gemin4 and the remaining strand is degraded (Lund et al., 2004). The ribonucleoproteins complexes are classified as RNA induced silencing complexes (RISC). RISC causes either target mRNA degradation or translation repression (**Figure 1.2**) (Schwarz and Zamore, 2002). Nelson et al., (2003) suggested the term RISC/miRNP endonuclease for the catalytic activity of RISC and miRNPs, since miRNAs with extensive sequence complementarity to their mRNA targets can trigger RNA interference (RNAi) (Andrea and Peter, 2004). If a miRNA is completely complementarily to its target mRNA, degradation of this mRNA will occur. If a miRNA is partially complementarily to its target mRNA, then translation of this mRNA will be inhibited.

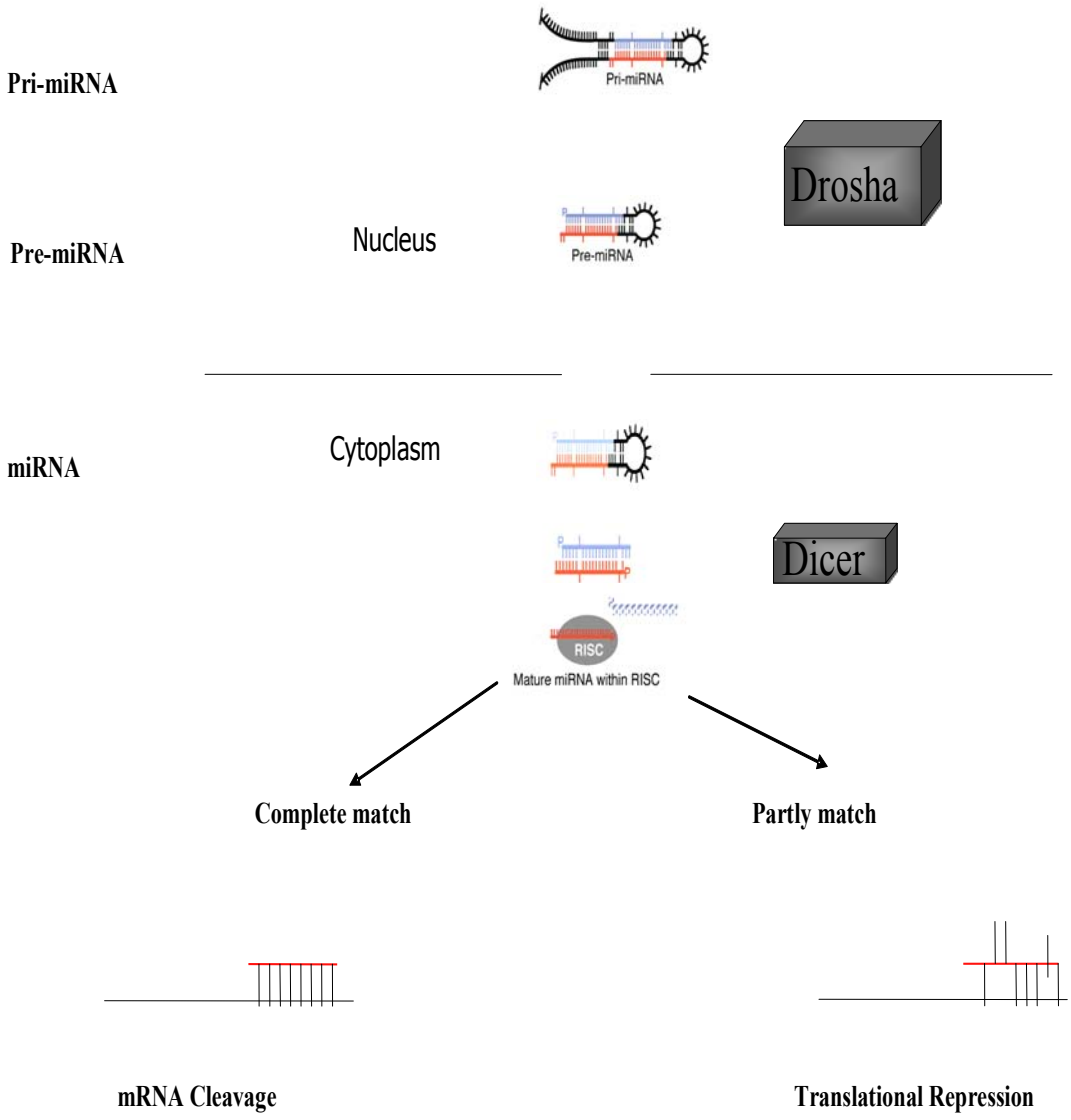

Figure 1.2. The development of the mature miRNAs (Bartel, 2004). Pri-miRNAs are cleaved by Drosha in the nucleus yielding smaller pre-miRNAs. pre-miRNAs are transported to the cytoplasm, where Dicer, a member of the RNAse III nuclease family, further processes them to yield mature miRNAs. miRNAs associate with RNA induced silencing complex (RISC), which affects the silencing of target mRNA molecules. If a miRNA is complementary to its target mRNA, degradation of this mRNA will occur. If a miRNA is partially complementary to its target mRNA, then translation of this mRNA will be inhibited.

1.3.1 miRNAs and Hematopoietic Cancers

miRNAs play important regulatory roles in animal and plant development at the post-transcriptional level. Many miRNAs cloned from mouse bone marrow cells are differentially regulated in various hematopoietic lineages, suggesting that they might influence hematopoietic lineage differentiation (Chena and Lodish, 2005). For example, mir-181, a miRNA specifically expressed in B-cells within mouse bone marrow, promotes B-cell differentiation when expressed in hematopoietic stem/progenitor cells. Some human miRNAs are associated with leukemias (Calin et al., 2002). mir-15a/mir-16 locus is frequently deleted or down-regulated in B-CLL patients and mir-142 is at a translocation site found in a case of aggressive B-cell leukemia. Collectively, these results indicate that miRNAs may be important regulators of mammalian hematopoiesis.

1.3.2 miRNAs and Cancer-Associated Genomic Aberrations

Many miRNAs are located at sites of cancer associated chromosomal aberrations. For example, the homozygous deletions of mir-15a and mir-16a cluster associate with aberration of deletion 13q14 in B-CLL (Calin et al., 2002). miRNAs that are located in the bands of chromosomal aberrations should predict low risk and high risk groups of B-CLL patients, Using microRNA microarrays, Calin et al. (2004) reported low level expression of mir-16-1 in leukemia harboring deletions at 13q14 (P=0.03, ANOVA), a chromosomal aberrations associated with prognosis of B-CLL. They also found a distinct miRNA signature composed of five differentially expressed miRNAs (mir-186,

mir-132, mir-16-1, mir-102, and mir-29C) that distinguished mutated and unmutated B-CLLs. mir-16-1 and mir-132, are associated with B-CLL chromosomal aberrations d13q14 and d17p13.3, respectively. Michael et al., found that the mir-143 and mir-145 miRNA cluster is located at fragile sites in genomic regions. The expression of this miRNA cluster is down-regulated in colon cancers (Michael et al., 2003).

1.4 Hypothesis

The main hypotheses of this thesis are that the miRNAs have significant differences in expression between mutated and unmutated B-CLL patient groups and that miRNAs located in genomic regions that are associated with B-CLL aberrations are useful for classifying B-CLL.

1.5 Specific Aims

We pursued the following specific aims to test our hypotheses. First, we identified miRNAs located at chromosomal aberrations sites associated with B-CLL. Second, we collected a set of peripheral blood samples from B-CLL patients and determined their mutational status. Third, we monitored the expression of a limited set of miRNAs in B-CLL patient material. Fourth, we correlated miRNA expression with mutational status.

CHAPTER 2

Materials and Methods

We performed the following tasks to compare the miRNA expression between mutated and unmutated B-CLL groups. In Task 1, we obtained B-CLL cells from patients. In Task 2, we obtained total RNA and cDNAs from B-CLL cells. In Task 3, we classified the mutational status of patients and divided them into mutated and unmutated B-CLL groups. In Task 4, we quantified miRNA expression levels. In Task 5, we analyzed the significant differences in miRNA expressions in the mutated and unmutated B-CLL groups. **Figure 2.1** shows a relationship among these tasks and highlights their inputs and outputs. The details of these tasks are presented below.

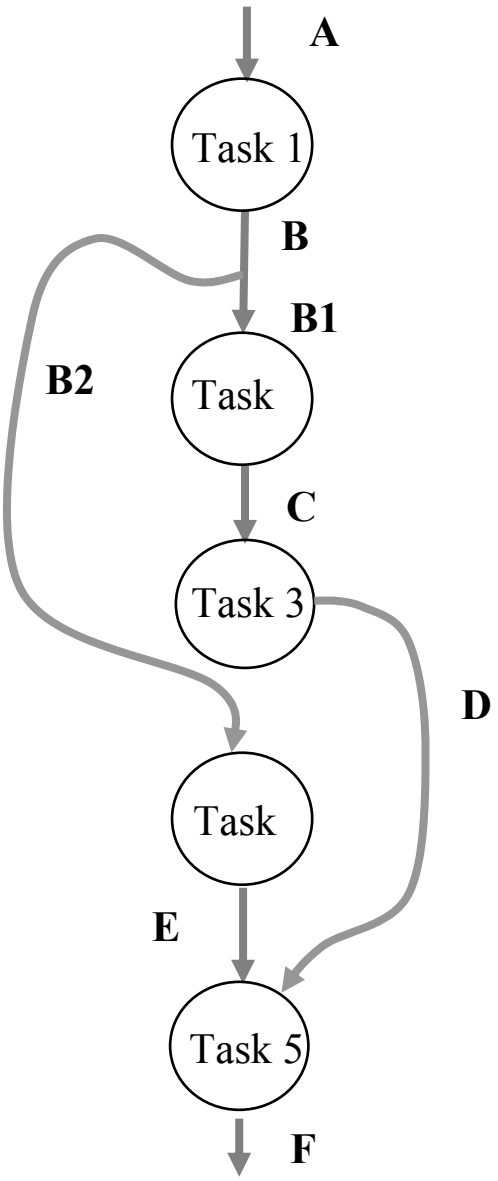

Figure 2.1. Relationship among task inputs and outputs. **A**: Peripheral Blood; **B**: Total B-cells. (B=B1+B2); **C**: IgV_H Gene Sequence; **D**: Mutational Status; **E**: miRNA; **F**: miRNA Expression Level.

2.1 Task 1: Collection of B-CLL Samples

2.1.1 Collection of Patient Material

We obtained thirty patient peripheral blood samples with the consent of patients diagnosed with B-CLL at the Royal University Hospital, Saskatoon, Saskatchewan. We chose patient samples for further analysis based on the percentage of white blood cell count (>5000/mm^3), and the percentage of B-cells in the white blood cells (> 95 %). We confirmed the B-CLL patient diagnosis using morphologic and immunophenotypic criteria. All B-CLL samples in this study were collected before any therapy was given.

2.1.2 Isolation of B-CLL cells

We used the following procedure to obtain B-CLL cells. We diluted anti-coagulated blood cells to a final volume of 30 ml in 1×phosphate-buffered saline (PBS). We layered the diluted anti-coagulated blood cells onto 9 ml of Ficoll Paque. We centrifuged the sample at 400 g for 20 minutes at room temperature. We discarded the upper layer of plasma and transferred the mononuclear layer to a 15 ml tube. We diluted the mononuclear layer with 1×PBS to a final volume of 10 ml. We centrifuged the diluted mononuclear cells at 400 g for 5 minutes at room temperature. We decanted the supernatant and re-suspend the cells in 10 ml of PBS. We removed 30 µl of cells and counted the cells by staining them with same volume of trypan blue. We washed the cells an additional two times using the same procedure described above. We re-suspended the final cell pellet at a concentration of 1×10^7 cell/ml in 10 % DMSO (Dimethyl Sulphoxide) and 40 % FCS (Fetal Calf Serum) and stored the cells at -80 °C.

2.2 Task 2: RNA Isolation, cDNA Synthesis, Amplification of IgV_H DNA, and DNA Sequencing

2.2.1 RNA Isolation

We used the following protocol to isolate total cellular RNAs from frozen B-CLL cells. We thawed the frozen cells, pelleted them by centrifugation, and re-suspended them in Trizol reagent (1×10^7 cells per ml Trizol reagent). We lysed the cells by vortexing them for 1 minute at room temperature. We added 0.2 ml of BCP (1-Bromo-3-Chloro-Propane) to each milliliter of lysed cells in Trizol reagent. We vortexed the mixture for 15 seconds and incubated it at room temperature for 2-3 minutes. We centrifuged the mixture for 15 minutes at 4 °C at 12000 g. We transferred the aqueous phase to a fresh tube and added 0.5 ml isopropyl alcohol per milliliter of Trizol reagent to precipitate RNA. We incubated the mixture (the aqueous phase with isopropyl) at room temperature for 10 minutes and centrifuged it for 10 minutes at 4 °C at 12000 g. We discarded the supernatant and washed the RNA pellet in 0.5 ml of 75 % ethanol. We vortexed the mixture and centrifuged it at 7500 g at 4 °C for 5 minutes. We discarded the ethanol supernatant and air-dried the pellet for 10-15 minutes. We dissolved the pellet in the RNase-free water (~ 25 µl). We determined the concentration of the isolated RNA using an UV spectrometer at an OD (Optical Density) of 260 and 280 nm.

2.2.2 cDNA Synthesis

We added 1 µg total RNA to a 20 µl reaction containing: 50 µmol of random hexamer primers (PE, Foster City, CA), 200 U of Super Script II, First Strand Buffer (50 mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM $MgCl_2$), 10 mM of DTT, 20 U of RNase inhibitor, and 0.5 mM of each dNTP (GIBCO BRL, Gaithersburg, MD). We incubated

the reaction mixture for 10 minutes at 21 °C and then for 30 minutes at 42 °C, using a GeneAmp System 2400 (PE, Foster City, CA), to produce the cDNA. We heated reaction mixture at 95 °C for 5 minutes to inactivate the Super Script II reverse transcriptase.

2.2.3 Polymerase Chain Reaction (PCR) Amplification of IgV_H cDNA

We amplified 1 µl of cDNA using GeneAmp System 2400 (PE, Foster City, CA) using the following 50 µl PCR reaction: PCR buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl), 1.5 mM $MgCl_2$ (for VHL1, VHL3, and VHL5) or 2.0 mM $MgCl_2$ (for VHL2, VHL4, and VHL6), 0.2 mM of each dNTP, and 1.25 U Hotstar polymerase (Hot Star Qiagen). We added 50 pmol of downstream primer corresponding to a consensus sequence (J_H, 5'- ACCTGAGGAGACGGTGACC-3') at the end of the J-region with 50 pmol of one of the following upstream primer corresponding to 1 of the 6 human V_H leader sequences:

VHL1, 5'- CCATGGACTGGACCTGGAGG- 3';

VHL2, 5'-ATGGACATACTTTGTTCCAGC-3';

VHL3, 5'-CCATGGAGTTTGGGCTGAGC-3';

VHL4, 5'-ATGAAACACCTGTGGTTCTT-3';

VHL5, 5'-ATGGGGTCAACCGCCATCCT-3';

VHL6, 5'-ATGTCTGTCT CCTTCCTCAT-3',

We performed one cycle of amplification by pre-activating "hotstar" DNA polymerase at 95 °C for 20 minutes, annealing the primers at 65 °C for 4 minutes, and extending the primers at 72°C for 1 minute. Then, we performed 40 cycles of amplification at the

following settings: 95 °C for 1 minute, 61 °C for 30 seconds, and 72 °C for 1 minute. Finally, we incubated the PCR reaction mixture at 72 °C for 6 minutes to produce the final PCR product. We analyzed the PCR product on a 2 % agarose gel stained with ethidium bromide and visualized the PCR product by exposing it with a UV Transilluminator (Bio-Rad Inc). We stored the PCR product at -20 °C.

2.2.4 Sequencing of IgV_H Regions

We determined the concentration of PCR products by running them together with a standard marker on a 2 % agarose gel stained with ethidium bromide. Based on the intensity of the PCR product relative to the standard pre-made marker, we determined the concentration of individual PCR-products. We sequenced the PCR-products (1 μg/each) using ABI PRISM: 310 Genetic Analyzer.

2.3 Task 3: Classification of IgV_H Mutational Status of Patient Samples

We classified B-CLL cells into mutated and unmutated groups based on the mutational status of the IgV_H region. We calculated the mutational status by comparing the patient IgV_H gene sequence to its corresponding germline sequence. If the IgV_H gene sequence differs from the germline sequence by > 2 %, then we classify the B-CLL as mutated. The reliability of this approach can be enhanced with an approach proposed by Lossos et al. called “antigen selection”. Details of these two approaches are given below (Lossos et al., 2000).

2.3.1 Comparing Patient IgV_H sequence to Germline Sequence

We aligned patient IgV_H gene sequences to immunoglobulin germline gene sequences that we obtained from the Immunoglobin Basic Local Alignment Search Tool (Ig BLAST) database (http://www.ncbi.nlm.nih.gov/igblast/) (National Center for Biotechnology Information, Bethesda, MD). Ig BLAST program identifies the IgV_H germline sequence that is most similar to the patient IgV_H sequence. When the patient IgV_H differs from the germline IgV_H sequence by more than 2 %, then B-CLL cells are classified as mutated. If the patient IgV_H differs from the germline IgV_H sequence by less than 2 %, then B-CLL cells are unmutated B-CLL.

2.3.2 Antigen Selection of B-CLL Cells

Lossos et al., (2000) developed an algorithm to calculate the probability that a B-Cell has undergone antigen selection (http://www-stat.stanford.edu/immunoglobin). The method used to predict whether antigen selection of Ig genes has occurred is based on assessing replacement (R) and silent (S) mutation in the complementary-determining regions (CDR) and framework regions (FR). Replacement mutation is a point mutation in a codon that causes a different amino acid to occur in the protein. Silent mutation is a point mutation in a codon, which due to the degeneracy of the genetic code, results in the same amino acid in the protein. The distribution of replacement (R) mutations and silent (S) mutations within the CDRs and framework regions (FRs) is used to determine whether the antigen selection occurs. If a more R mutations than S mutations occur in the complementary-determining region (CDR) than in the framework region (FR), then

the antigen selection is predicted to have occurred. During antigen selection, somatic mutation occurs, which is predicted to give rise to the mutated B-CLL group.

To calculate the probability that a B-CLL cell has undergone antigen selection using Lossos' algorithm, the following information is needed. First the germline sequence for the IgV_H region is obtained from http://www.ncbi.nlm.nih.gov/igblast. Second, the germline sequence is divided into FR1, FR2, FR3, CDR1, and CDR2 regions. Third, the numbers of R and S mutations in these regions are calculated. This information is used to obtain P-values for whether antigen selection has occurred.

2.4 Task 4: miRNA Isolation, miRNA Probe Construction, and miRNA Detection

2.4.1 miRNA Isolation

We used 5×10^7 frozen patient B-CLL cells to isolate miRNA using *mir*Vana™ miRNA Isolation Kit 1560 (Ambion). The specific details of the isolation protocol are described in the *mir*Vana™ miRNA Isolation Kit 1560 manual. The basic concepts behind the miRNA isolation protocol are as follows. The first step in miRNA isolation is to disrupt B-CLL cells in denaturing lysis buffer. The second step is to subject the samples to an Acid-Phenol:Chloroform extraction, which removes most of the protein and DNA from the sample (Chirgwin et al., 1979). The third step is to enrich the sample for small RNAs. To do this, large RNAs are removed from the sample by immobilizing them on a glass filter. Large RNAs are bound to the glass filter when they are applied to the glass filter in a solution containing a low concentration of ethanol (25 %).

Smaller RNAs do not bind to the glass filter at this ethanol concentration. Small RNAs can be isolated by the glass filter by increasing the ethanol concentration to 55 %. The detailed procedure for the isolation is described as follows (see also **Figure 2.2**)

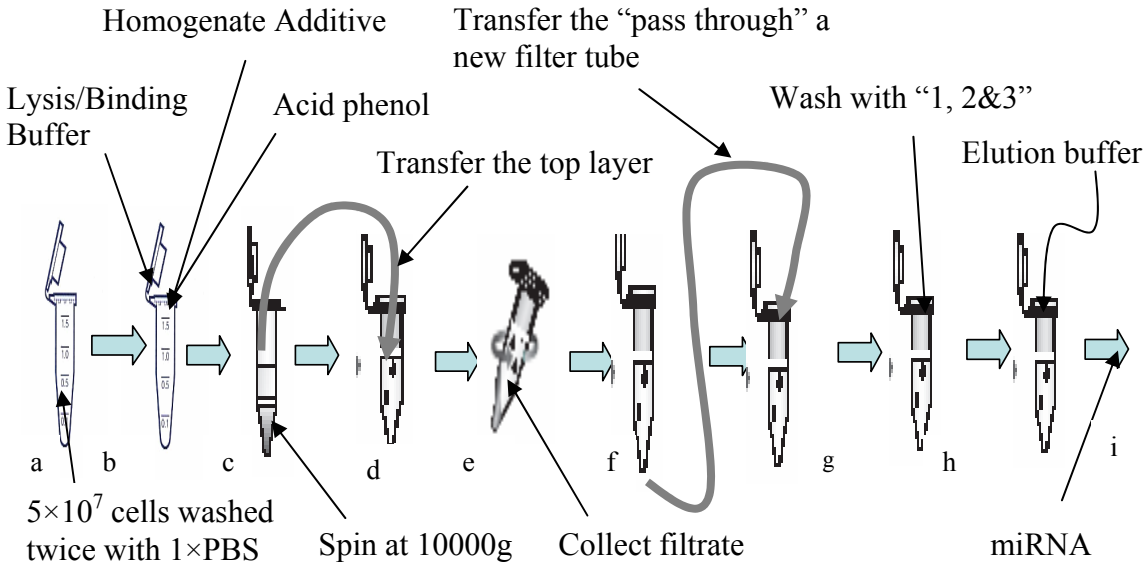

Figure 2.2. The isolation process for miRNA. After the cells were washed and pelleted (a), the first step is to disrupt B-CLL cells in denaturing lysis/binding buffer which is provide in kits, then the samples were subjected to an Acid-Phenol:Chloroform extraction to removes most of the protein and DNA from the sample (b). The aqueous phase was transferred to a fresh tube and the volume of the transferred aqueous phase was recorded (c). 1/3 volume of the 100% ethanol was added to the RNA product (d), then centrifuge the mixture to pass through a filter cartridge (e), first filter cartridge, and collected the filtrate, and record the volume of filtrate (f). 2/3 volume of the 100% ethanol was added to the filtrate, then transfer the mixture into a second filter cartridge and centrifuged (g). After washed with wash solution 1, 2 & 3 that are provide in kits (h), elute the small RNA from the filter (i).

We measured the small RNA concentration using a smartspec^{TM} 3000 UV spectrometer (Bio-Rad Inc). We analyzed 0.25 µg of small RNAs on a 15 % polyacrylamide gel (together with 0.25 µg mir-16) and visualized the RNA by staining with cyberGold (Molecular Probes Inc) and monitoring fluorescence using an UV Transilluminator (Bio-Rad Inc). We normalized the small RNA concentrations from different patient

samples using a common RNA (usually the 5.8S ribosomal RNA). We chose one patient sample as the standard to normalize all patient samples against. The recovered small RNAs are stored it at –80 °C until further analysis.

2.4.2 Solution Hybridization Assay to Detect miRNA

We used the solution hybridization assay developed by Ambion (mirVana miRNA detection kit) to detect specific miRNAs presents in the patient samples. The specific details of the isolation protocol are described in the (mirVana miRNA detection kit) manual. The basic concepts behind the miRNA isolation protocol are summarized in **Figure 2.3.** miRNAs are detected in small RNA patient samples by hybridizing specific miRNAs to a radiolabeled antisense RNA probe. Unhybridized probes and miRNA are then removed from the sample by a ribonuclease digestion. Double-stranded radiolabeled probes that are protected RNase digestion are purified and detected using a denaturing polyacrylamide gel.

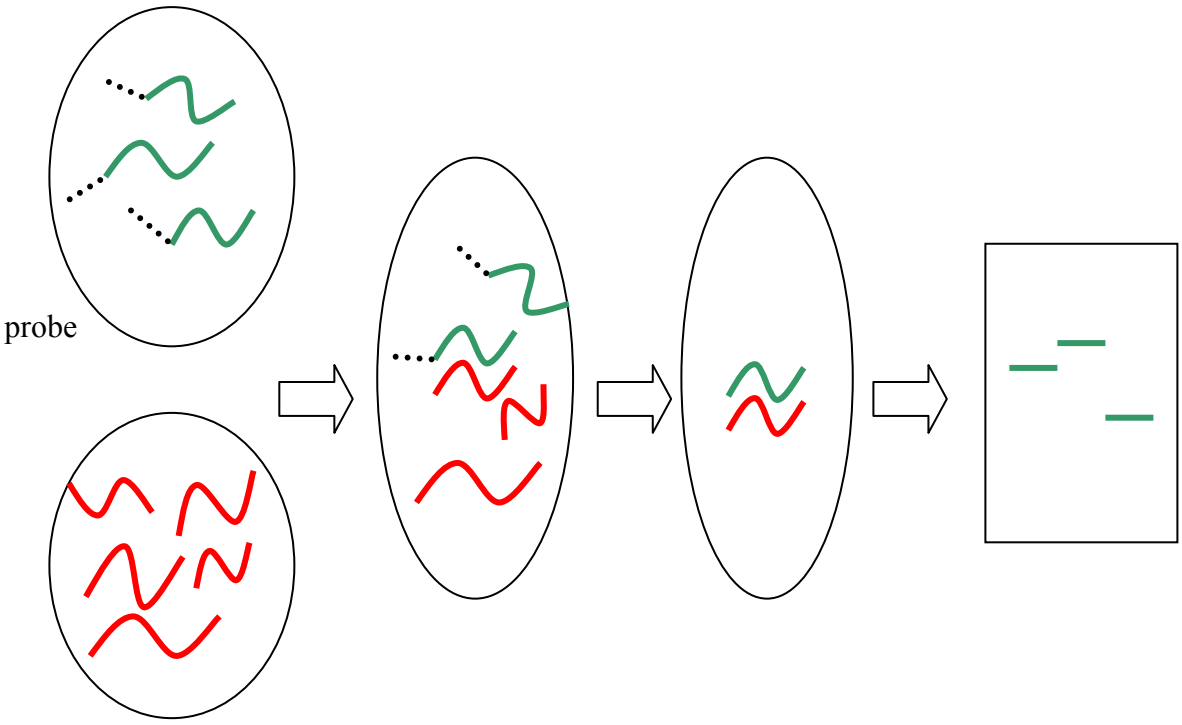

Figure 2.3. Solution Hybridization Assay. miRNAs in small RNA patient samples are detected by hybridizing with a specific radiolabeled antisense RNA probe. (a) miRNA probe is mixed with the sample RNAs. (b) The miRNAs are hybridized to their complementary probes. (c) The mixture is treated with RNase to digest unhybridized probe. (d) Double-stranded radiolabeled probes that are protected against RNase digestion are detected using a denaturing polyacrylamide gel.

The solution hybridization assay is divided into three steps: probe construction, miRNA RNase protection assay, miRNA detection.

2.4.3 Probe Construction

We prepared high specific-activity ^{32}P-labeled antisense RNA probes using *in vitro* transcription. The *in vitro* transcription reaction is consists of four steps, which are summarized in **Figure 2.3**. In the first step, DNA oligonucleotide nucleotide templates

are designed to construct antisense probes for specific miRNAs. At the 5' end, the DNA template contains a sequence that is identical to the miRNA that will be detected. At the 3' end, the DNA template contains an eight base sequence complementary to the 3' end of the T7 Promoter primer. The DNA templates for the eight miRNAs used in this study are listed in **Table 2.1**. Several of the DNA templates contain four T bases between the miRNA sequence and the T7 complementary region. The extras Ts allow several probes to be assayed simultaneously. The addition of T residues generates an antisense RNA probe that contains a stretch of A residues that cannot be cleaved by RNases A nor T1. (Ausubel et al., 2004). The addition of 4 Ts residues to the DNA oligonucleotide nucleotide template sequence is sufficient for size resolution from the 0 T DNA oligonucleotide nucleotide template on the PAGE gels used for multi-probe solution hybridization assays.

Table 2.1. Eight selected miRNAs: Their mature sequences and DNA oligonucleotide templates

miRNA	**miRNA Sequence (Mature) and DNA oligonucleotide for T7 miRNA Probe (+4Ts and 0T)**
hsa-miR-34**b**	14AGGCAGUGUCAUUAGCUGAUUG35 AGGCAGTGTCATTAGCTGATTG*TTTT***CCTGTCTC**
hsa-miR-125b	15UCCCUGAGACCCUAACUUGUGA36 TCCCTGAGACCCTAACTTGTGA**CCTGTCTC**
hsa-miR-16-1	14UAGCAGCACGUAAAUAUUGGCG35 TAGCAGCACGTAAATATTGGCG**CCTGTCTC**
hsa-miR-196-2	25UAGGUAGUUUCAUGUUGUUGGG46 TAGGTAGTTTCATGTTGTTGGG*TTTT***CCTGTCTC**
hsa-miR-26a-2	14UUCAAGUAAUCCAGGAUAGGCU35 TTCAAGTAATCCAGGATAGGCT**CCTGTCTC**
hsa-let-7i	6UGAGGUAGUAGUUUGUGCU24 TGAGGTAGTAGTTTGTGC*TTTTT***CCTGTCTC**
hsa-miR-181c	27AACAUUCAACCUGUCGGUGAGU48 AACATTCAACCTGTCGGTGAG*TTTTT***CCTGTCTC**
miR-125a	15UCCCUGAGACCCUUUAACCUGUG37 TCCCTGAGACCCTTTAACCTGTG*TTTT***CCTGTCTC**

In the first step, mature miRNA sequences were obtained from 'Sanger Rfam' database (http://www.sanger.ac.uk/Software/Rfam/mirna/index.shtml).

In the second step, a double-stranded transcription template is prepared. The oligonucleotide nucleotide template is annealed to the T7 Promoter Primer (**Figure 2.4b**). The T7 promoter primer and the oligonucleotide nucleotide template are extended using Exo-Klenow DNA polymerase (**Figure 2.4c**).

In the third step, the double-stranded DNA template is transcribed using T7 phage RNA polymerase (**Figure 2.4d**). The RNA polymerase binds to the double-stranded template and separates the two strands of DNA. The RNA polymerase uses the 3' to 5' strand as a template to synthesize a complementary 5' to 3' RNA transcript. The RNA transcript was radiolabeled by incorporating [α-^{32}P] UTP (Amersham Inc) *in vitro* transcription reaction. The DNA template in the reaction is removed using DNase I. We purified the RNA probes made by *in vitro* transcription to remove transcripts that are shorter than the full-length probe because these shorter products may cause "background" smears and/or spurious protected bands in the assay. We purified the RNA transcripts using a denaturing polyacrylamide gel. We determined the yield and the specific activity of radiolabeled transcripts measuring and comparing the total amount of radiolabel present in the reaction mixture using a scintillation counter (BECKMAN Inc).

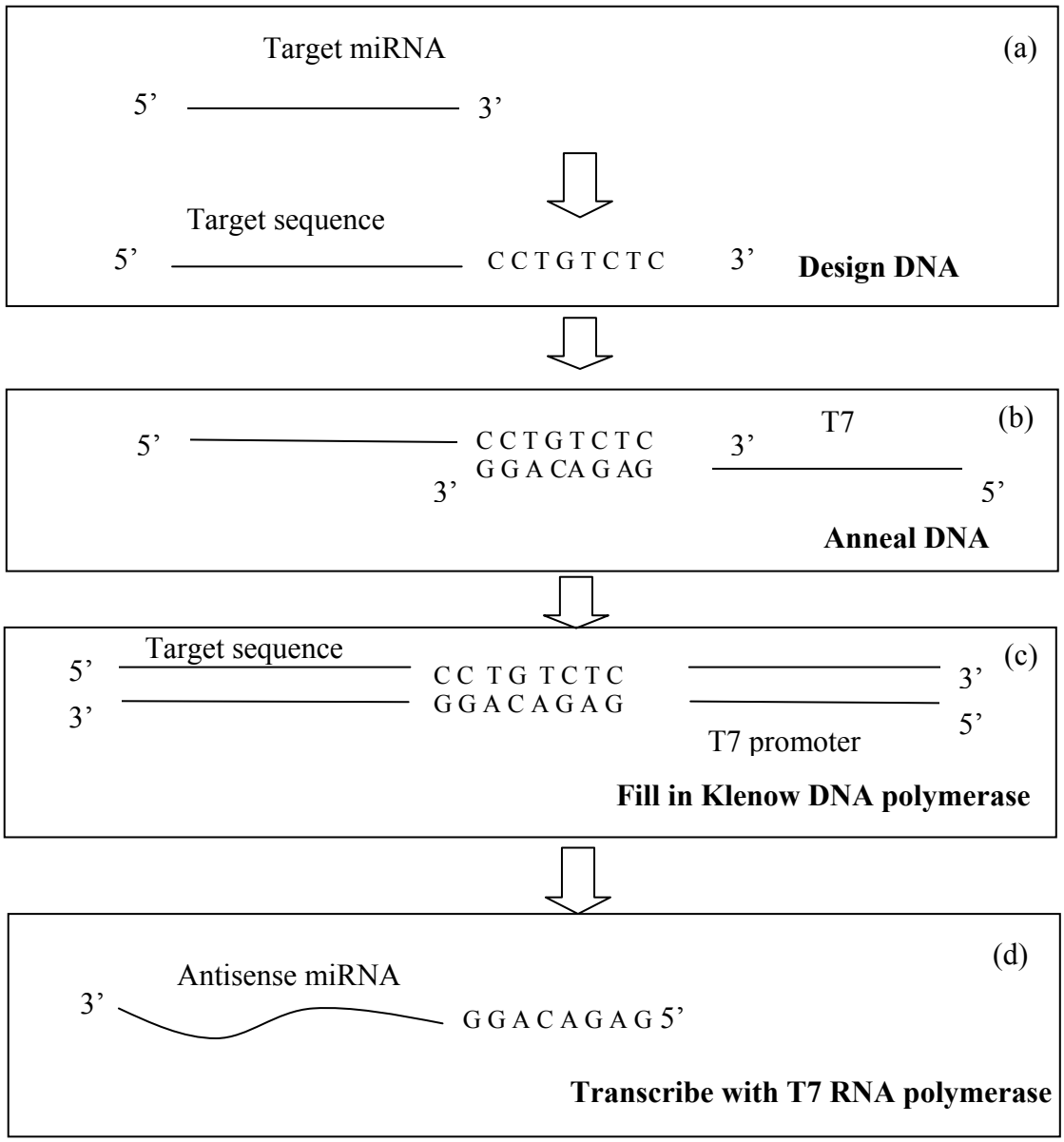

Figure 2.4. miRNA probe construction. (a) The DNA oligonucleotide nucleotide has the same sequence as its target miRNA, except the U residue is replaced by T residue. An eight base sequence which is complementary to the 3' end of the T7 Promoter Primer was added to the 3’end of oligonucleotide . (b) T7 promoter primer was annealed to the oligonucleotide nucleotide. (c) Double-stranded DNA template was synthesized using Klenow DNA polymerase. (d) miRNA probe was generated using T7 RNA polymerase. Two extra Gs are added during the transcription reaction.

miRNA probe construction for mir-16 and mir-16^{+4} are outlined below.

mir-16 Sequence: 5'-UAGCAGCACGUAAAUAUUGGCG-3'

DNA template for mir-16: 5'-TAGCAGCACGTAAATATTGGCG**CCTGTCTC**-3'

Mir-16 probe: 3'-AUCGUCGUGCAUUUAUAACCGCGGACAGAGGG-5'

Product from solution hybridization assay: 3'-AUCGUCGUGCAUUUAUAACCGC-5'

mir-16^{+4} Sequence: 5'-UAGCAGCACGUAAAUAUUGGCG-3'

DNA template for mir-16^{+4}:

5'-TAGCAGCACGTAAATATTGGCGTTTT**CCTGTCTC**-3'

mir-16^{+4} Probe: 3'-AUCGUCGUGCAUUUAUAACCGCAAAAGGACAGAGGG-5'

Product from solution hybridization assay:

3'-AUCGUCGUGCAUUUAUAACCGCAAAA-5'

2.4.4 miRNA Detection: miRNA RNase Protection Assay

Specific antisense RNA probes are used to detect miRNAs present in small RNAs samples from patient material. If a specific miRNA probes is present in the sample, then it will protect the RNA probe from degradation by RNase. The amount of protected probe is directly proportional to the amount of miRNA in the sample. Typically, we use 0.5 to 5 μg of small RNA sample for the RNase protection assay. We add ^{32}P-labeled RNA probe in five-fold molar excess over the target miRNA. In most cases 5×10^{4} cpm (0.023 μCi) of highly specific activity probes meet this requirement.

In addition to the small RNA patient samples, we use two miRNA controls and one standard marker for each probe in the experiment. The first control contains miRNA probe and no miRNA or RNase and it is used to assess probe quality and recovery (negative control). The second control contains miRNA probe and no miRNA and it is used to assess non-specific background signals and verify that probe digestion is complete (positive control). The standard marker contains miRNA probe and standard miRNA.

The total amount of RNA in each sample and control should be roughly equal (+/– 10 %) in order to ensure that the RNase digestion conditions are similar in each. If the amounts of sample RNAs vary, we used yeast RNA to adjust the total amount of RNA in each hybridization reaction, plus controls up to 5 μg.

The antisense RNA is hybridized to the miRNAs by heating it to 95 °C and slowly cooling to 52 °C over a period of 2 hours. This step denatures the RNA, specifically reducing the effect of some secondary structure that may be formed during hybridization. The hybridized RNA samples are treated with RNase A and T1 to digest RNAs that do not form complexes with the antisense RNA probe. The protected RNA probes are precipitated and detected using a denaturing polyacrylamide gel. The amount of protected ^{32}-P antisense RNA probe protected was quantified using a phosphorimager (Bio Rad Inc).

2.5 Task 5: Statistical Analysis of miRNA Expressions

We determined the intensities of miRNA bands from the solution hybridization detection gels using a phosphoimager (Bio Rad Inc). We used the significance test facility in SAS (Statistical Analysis Software) software (SAS Institute, Cary NC) to evaluate whether there are significant differences in the miRNA expression between the mutated and unmutated B-CLL groups.

2.5.1 Statistical Problem Formulation

The evaluation problem can be defined as a significance test problem in statistical data analysis. For the significance test for two groups (the mutated and unmutated B-CLL groups), we used the t-test (Appendix A). Assume that for a particular miRNA (e.g., mir-26a-2) we have its expressions (Intensities from phospoimager) in the two groups as follows:

(i) In the mutated group: 4271839.8, 5154553.9, 2639191, 3606797, 3216792.2, 3688059.3, 1894745.6, 2793187.3

(ii) In the unmutated group: 1421870.9, 1568594.1, 1803838.1, 2179905.6

In this case, group (i) has 8 elements, and group (ii) has 4 elements. We denote μ_1 for the mean for group (i) population and μ_2 for the mean for group (ii) population.

We assumed that the miRNA expression followed a normal distribution and two groups were mutually exclusive. Using these assumptions, our problem satisfied the assumptions for the t-test (Appendix A). Note that the miRNA expressions of the two

groups may not have an equal variance. We used the Levene test to evaluate the equal variance or unequal variance (Appendix A). Note that the t-test uses different equations depending on whether there is equal or unequal variance (see Appendix A for details). Furthermore, in our case, we chose the significance level $\alpha = 0.05$. The t-test was a two-tailed test.

2.5.2 Procedure For Using SAS

We programmed the problem in conformity with the input file format of SAS for the t-test; see below (for the example of mir-26a-2).

```
data first;
input Rep species$ intensity;
Cards;
1     M     4271839.8
1     M     5154553.9
1     M     2639191
1     M     3606797
1     M     3216792.2
1     M     3688059.3
1     M     1894745.6
1     M     2793187.3
2     UM    1421870.9
2     UM    1568594.1
2     UM    1803838.1
2     UM    2179905.6
;
proc ttest;
     class species;
     var   intensity;
          title 'T-test using SAS ttest';
run;
```

We then run the SAS program and obtained the output file below.

T-test using SAS ttest 10:56 Friday, August 13, 2004 1
The TTEST Procedure
Statistics

Variable	patients	N	Mean	Mean	Mean	Std Dev	Std Dev	Std Dev	Std Err
intensity	M	8	2.56E6	3.41E6	4.26E6	671027	1.01E6	2.07E6	358822
intensity	UM	4	1.22E6	1.74E6	2.27E6	187350	330720	1.23E6	165360
intensity	Diff (1-2)		479930	1.66E6	2.85E6	606650	868235	1.52E6	531683

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
intensity	Pooled	**Equal**	10	3.13	**0.0107**
intensity	Satterthwaite	Unequal	9.31	4.21	0.0021

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
intensity	Folded F	7	3	9.42	**0.0924**

We first read information of {Pr > F} from the third box above. If {Pr > F} is greater than $\alpha = 0.05$ (this is the case for the example), we conclude that two groups have unequal variance (this is the case for the example). Then we read from the second box information of {Pr > |t|} according to the equal variance or unequal variance case. If {Pr > |t|}< α=0.05, then we conclude that two groups have significant difference in terms of their means. In this example, we found {Pr > |t|} = 0.0107 that is less than $\alpha = 0.05$; So the two groups have significant difference in terms of their means of miRNA expressions.

CHAPTER 3

Results and Discussion

In total, we characterized thirty B-CLL patient samples in this study. The results include (i) the classification of samples into mutated B-CLL and unmutated B-CLL groups, and (ii) the characterization of miRNA expressions in these two patient groups. This chapter presents these results with discussions. Section 3.1 presents result followed by a discussion in section 3.2.

3.1 Results

3.1.1 Patient B-CLL IgV_H Mutation Status

We obtained thirty patient peripheral blood samples (B-CLL cells) for the experiments conducted in this study. We isolated mononuclear white blood cells using Ficoll Paque. These mononuclear white blood cells consist of more than 95% B-cells. We isolated

total RNA from the B-CLL cells using Trizol. We converted poly A tailed RNA to cDNA using reverse transcriptase. We used PCR to amplify the IgV_H region from the cDNA from each patient using six primer sets. Each primer set contains a unique upstream primer corresponding to one of the six human V_H leader sequences and a constant J-region primer. B-CLL is a monoclonal disease and all B-CLL cells arise from a common cell. Therefore, B-CLL cells from each patient have identical rearranged variable regions.

Figure 3.1 shows PCR products for patients (PT) 1 to 4 (PCR products for the remainder of patients can be found in Appendix B). **Table 3.1** summarizes the primers sets that gave PCR products for all 30 patients.

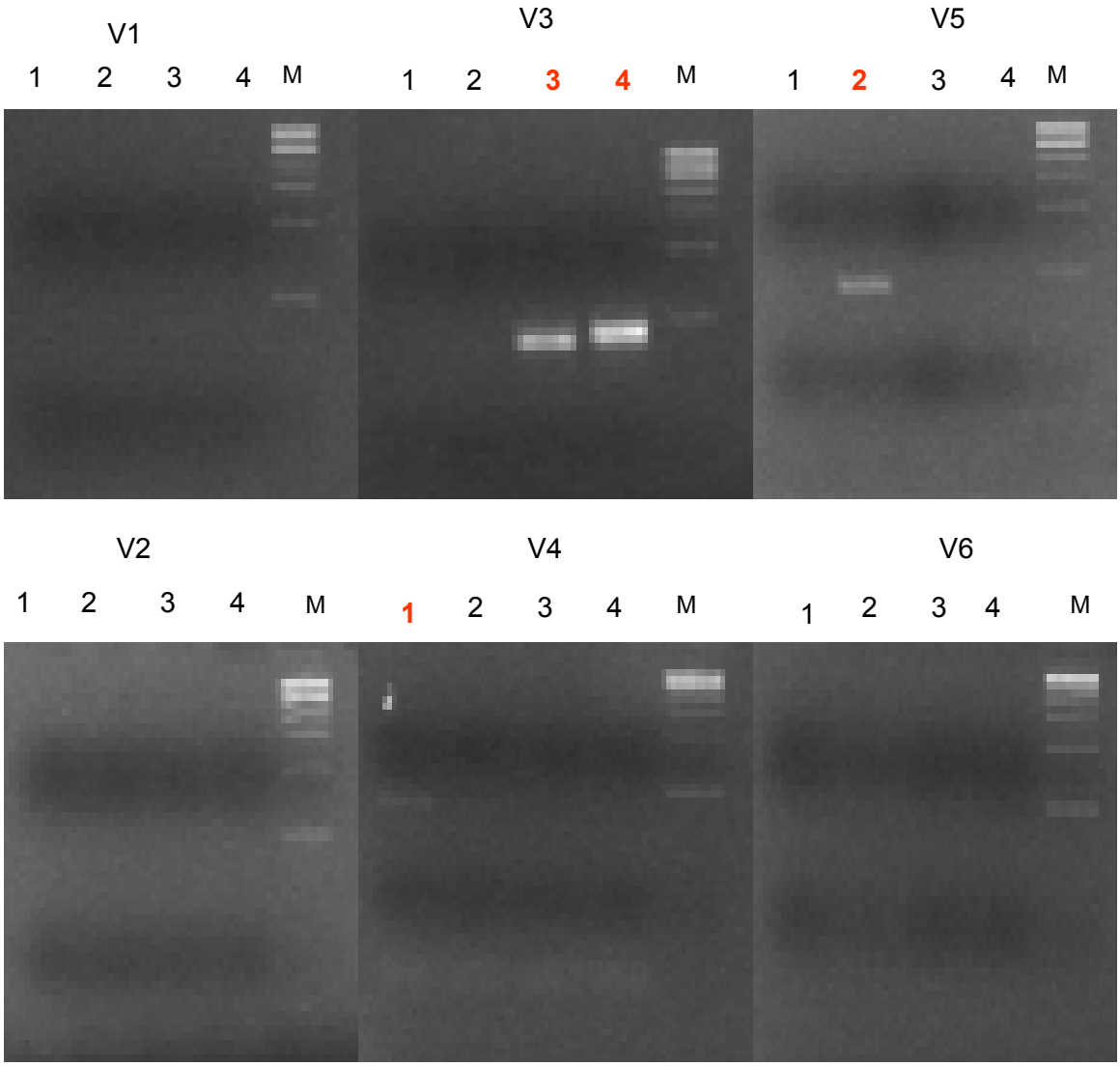

Figure 3.1. RT-PCR products for patients (PT) 1-4. PCR produces are separated on a 2 % agarose gel and stained with ethidium bromide. V1 to V6 correspond to the human V_H leader sequences primers. M represents a DNA marker.

Table 3.1. Summary of the IgV_H region RT-PCR for thirty patients (PT). (+) indicates primer that gave positive RT-PCR products.

	V1	V2	V3	V4	V5	V6		V1	V2	V3	V4	V5	V6
PT1				+			PT16					+	
PT2					+		PT17			+			
PT3			+				PT18				+		
PT4			+				PT19				+		
PT5			+				PT20			+			
PT6					+		PT21	+					
PT7					+		PT22				+		
PT8			+				PT23			+			
PT9	+						PT24	+					
PT10			+				PT25				+		
PT11				+			PT26			+			
PT12	+						PT27			+			
PT13	+						PT28			+			
PT14					+		PT29			+			
PT15			+				PT30	+					

To determine the mutational status of the patients, we sequenced the IgV_H region RT-PCR products using an ABI PRISM 310 Genetic Analyzer (DNA sequencer). A summary of the IgV_H region sequences for the thirty patients is provided in Appendix C. To illustrate the methodology used to calculate the B-CLL patient mutational status, we will use PT1 as an example.

The sequence of PT1 is:

CCAGATGGGTCTGTCCAGGTGCAGCTCAGCAGTGGGGCNCAGGACTGTTGA

AGCCTTCGGAGACCCTGTCCCTCACCTGCGCTGTCTATGGTGGGTCCTTCAG

TGATTACTACTGGACCTGGGTCCGCCAGCCCCCAGGGGGGGGTCTGGAGTG

GATTGGGNAAATCCATCACATTGGAGGTGCCAAATACAATCCGTCCCTCAA

GAGTCCAGTCACCATATCAATGGACACGTCCAAGAGCCAGTTCTCCCTGAG

ACTGATTTCTGTGACCGCCGCGGACACGGCTGTATATTACTGTGCGAGACTC
CCTCCGGAGTGGTTATTTCCTTTTGACTCCTGGGGCCCGGGGAACCCCGGTC
ACCGTCTCCTCAGGTAC

To determine the germline sequence that PT1 IgV_H region is derived from, we matched the patient's sequence to a database of germline sequences using Ig BLAST (http://www.ncbi.nlm.nih.gov/igblast). For PT1, the corresponding germline sequence is VH4-34 (V_H 4 gene family and particularly of V_H 4-34 gene of immunoglobulin heavy-chain variable region). We compared CDR1, CDR2, FWR1, FWR2, and FWR3 between PT1 and VH4-34 to determine the nucleotide differences in the DNA sequence. We then converted the DNA sequence to a protein sequence. We identified S and R mutations. S mutations are changes in the DNA sequence that do not change the amino acid sequence. R mutations are changes in the DNA sequences that change the amino acid sequence.

Calculation of PT1 Mutational Status

CDR1:

Comparison of CDR1 Region DNA Sequence

PT1: G**A**TTACTACTGGA**C**C

VH4-34: GGTTACTACTGGAGC

Nucleotide Mutations: 2

Comparison of CDR1 Region Amino Acid Sequence

PT1: DYYWT

VHL4-34: GYYWS

Amino Acid mutations: 2

Summary

R Mutations: 2

S Mutations: 0

Sequence length: 15 nucleotides

CDR2:

Comparison of CDR2 Region DNA Sequence

PT1: GAAATCCATCACATTGGAGGTGCCAAATACAATCCGTCCCTCAAGAGT

VHL4-34: GAAATCAATCATAGTGGAAGCACCAACTACAACCCGTCCCTCAAGAGT

Nucleotide Mutations: 8

Comparison of CDR2 Region Amino Acid Sequence

PT1: EIHHIGGAKYNPSLKS

VHL4-34: EINHSGSTNYNPSLKS

Amino Acid mutations: 5

Summary

R Mutations: 5

S Mutations: 3

Sequence length: 48 nucleotides

FWR1:

Comparison of FWR1 Region DNA Sequence

PT1:

CAGGTGCAGCTACAGCAGTGGGGCGCAGGACTGTTGAAGCCTTCGGAGACCCTGTCCCT

CACCTGCGCTGTCTATGGTGGGTCCTTC

VHL4-34:

CAGGTGCAGCTACAGCAGTGGGGCGCAGGACTGTTGAAGCCTTCGGAGACCCTGTCCCT

CACCTGCGCTGTCTATGGTGGGTCCTTC

Nucleotide Mutations: 0

Comparison of FWR1 Region Amino Acid Sequence

PT1: QVQLQQWGAGLLKPSETLSLTCAVYGGSF

VHL4-34: QVQLQQWGAGLLKPSETLSLTCAVYGGSF

Amino Acid mutations: 0

Summary

R Mutations: 0

S Mutations: 0

Sequence length: 87 nucleotides

FWR2:

Comparison of FWR2 Region DNA Sequence

PT1: TGG**G**TCCGCCAGCCCCCAGGG**GG**GGG**T**CTGGAGTGGATTGGG

VHL4-34: TGGATCCGCCAGCCCCCAGGGAAGGGGCTGGAGTGGATTGGG

Nucleotide Mutations: 4

Comparison of FWR2 Region Amino Acid Sequence

PT1: W**V**RQPPG**G**GLEWIG

VHL4-34: WIRQPPGKGLEWIG

Amino Acid mutations: 2

Summary

R Mutations: 2

S Mutations: 2

Sequence length: 42 nucleotides

FWR3:

Comparison of FWR3 Region DNA Sequence

PT1:

C**C**AGTCACCATATCA**A**T**G**GACACGTCCAAGA**G**CCAGTTCTCCCTGA**GA**CTGA**TT**TCTGT

GACCGCCGCGGACACGGCTGT**A**TATTACTGTGCGAGA

VHL4-34:

CGAGTCACCATATCAGTAGACACGTCCAAGAACCAGTTCTCCCTGAAGCTGAGCTCTGT

GACCGCCGCGGACACGGCTGTGTATTACTGTGCGAGA

Nucleotide Mutations: 9

Comparison of FWR3 Region Amino Acid Sequence

PT1: **P**VTIS**M**DTSK**S**QFSL**R**L**I**SVTAADTAVYYCAR

VHL4-34: RVTISVDTSKNQFSLKLSSVTAADTAVYYCAR

Amino Acid mutations: 5

Summary

R Mutations: 5

S Mutations: 4

Sequence length: 96 nucleotides

We calculated the mutational status by determining the percentage of nucleotides that are mutated in the CDR1, CDR2, FWR1, FWR2, and FWR3. For PT1 the total number of mutations is 23 and the total sequence length is 288. The total mismatch frequency is 23/288 or 7.98 %. In this case, there is 7.98 % (>2 %) difference in the sequence of PT1 from its germline sequence. Therefore, the B-CLL cells of PT1 went through the germinal centre, and its IGV_H gene underwent somatic hypermutation.

We also used the R and S mutations to calculate the probability that a B-CLL cell has undergone antigen selection using Lossos algorithm (Lossos et al., 2000), which is

available at the public website (http://www-stat.stanford.edu/immunoglobin). This algorithm calculates a P-value for whether antigen selection has occurred.

Table 3.2 summarizes the results of the analysis. A complete summary of patient information is in Appendix D. The sequence data from patient peripheral blood samples for PT11 and PT23 are missing CDR and FWR in their sequence. In **Table 3.2** not all sample cells obtained P-values which are less than 0.05. Mutation frequency and antigen selection probability calculation do not always classify patients into the same B-CLL subtypes; see **Table 3.2**. This might be quite natural as both methods have a subjective or uncertain component and are used in a different context; in the case of method (1) the threshold 2 % is a subjective component, while in the case of method (2) two P-values (with the confidence level 0.05) implies another subjective component. For our classification, we use only the percentage of mutations to classify B-CLL patients as mutated or unmutated.

Table 3.2. Upstream primer, P-value, and mutational status of B-CLL patients. ID indicates the percentage similarity of patient IgV_H gene to its germline sequence. P means the probability of antigen selection occurred. M: Mutated, UM: Unmutated.

				CDR	FWR	
Sample	**V_H family**	**VH gene**	**% ID**	**P**	**P**	**Pattern**
PT1	VH4	V4-34	92.02%	0.0698	0.0056	M
PT2	VH5	V5-51	95.83%	0.05143	0.04806	M
PT3	VH3	V3-74	99.40%	0.1833	0.4291	UM
PT4	VH3	V3-33	100%	0	1	UM
PT5	VH3	V3-23	94.60%	0.0902	0.001	M
PT6	VH5	V5-51	91.90%	0.0257	0.1538	M
PT7	VH5	V5-51	99%	0.1767	0.4082	UM
PT8	VH3	V3-74	92.52%	0.1479	0.0279	M
PT9	VH1	V1-69	99.67%	0.5876	0.2101	UM
PT10	VH3	V3-11	91.10%	0.1022	0.0581	M
PT11	VH4	-	-	-	-	-
PT12	VH1	VH1-2	100%	0	1	UM
PT13	VH1	VH1-69	98.70%	0.0098	0.01536	UM
PT14	VH5	VH5-51	97.62%	0.1738	0.32905	M
PT15	VH3	VH3-30	92.80%	0.04432	0.01449	M
PT16	VH5	VH5-51	95.30%	0.03047	0.04111	M
PT 17	VH3	VH3-11	99.32%	0.18247	0.09079	UM
PT 18	VH4	VH4-34	94.40%	0.03408	0.00609	M
PT 19	VH4	VH4-39	93.70%	0.02146	0	M
PT 20	VH3	VH3-33	96.94%	0.03843	0.22674	M
PT 21	VH1	VH1-2	95.92%	0.0013	0.01329	M
PT 22	VH4	VH4-28	91.31%	0.02418	0.02874	M
PT 23	VH1	-	-	-	-	-
PT 24	VH1	VH1-3	97.96%	0.19182	0.02691	M
PT 25	VH4	VH4-34	95.50%	0.55921	0.03206	M
PT 26	VH3	VH3-9	96.40%	0.02584	0.09563	M
PT 27	VH3	VH3-48	96.02%	0.07581	0.05312	M
PT 28	VH3	VH3-11	95.92%	0.001	0.05378	M
PT 29	VH3	VH3-33	95.24%	0.166	0.13545	M
PT 30	VH1	VH1-2	100.00%	0	1	UM

3.1.2 miRNA Isolation

We isolated miRNA form patient B-CLL cells using the *mir*Vana™ miRNA Isolation Kit 1560 (Ambion). We determined the RNA concentrations by measuring the Absorbance at 260 nm. We determined the quality of the isolated RNA by separating 0.25 µg of small RNA on a 15 % polyacrylamide gel and staining the RNA with cybergold (Molecular Probes Inc). **Figure 3.4** shows an example of RNA isolated from B-CLL cells for PT1-9. The remainder of the patients RNA is in Appendix E. We normalized the small RNA concentrations from different patient samples using a common RNA (usually the 5.8S ribosomal RNA). **Table 3.3** summarizes the normalized small RNA concentrations for each patient sample.

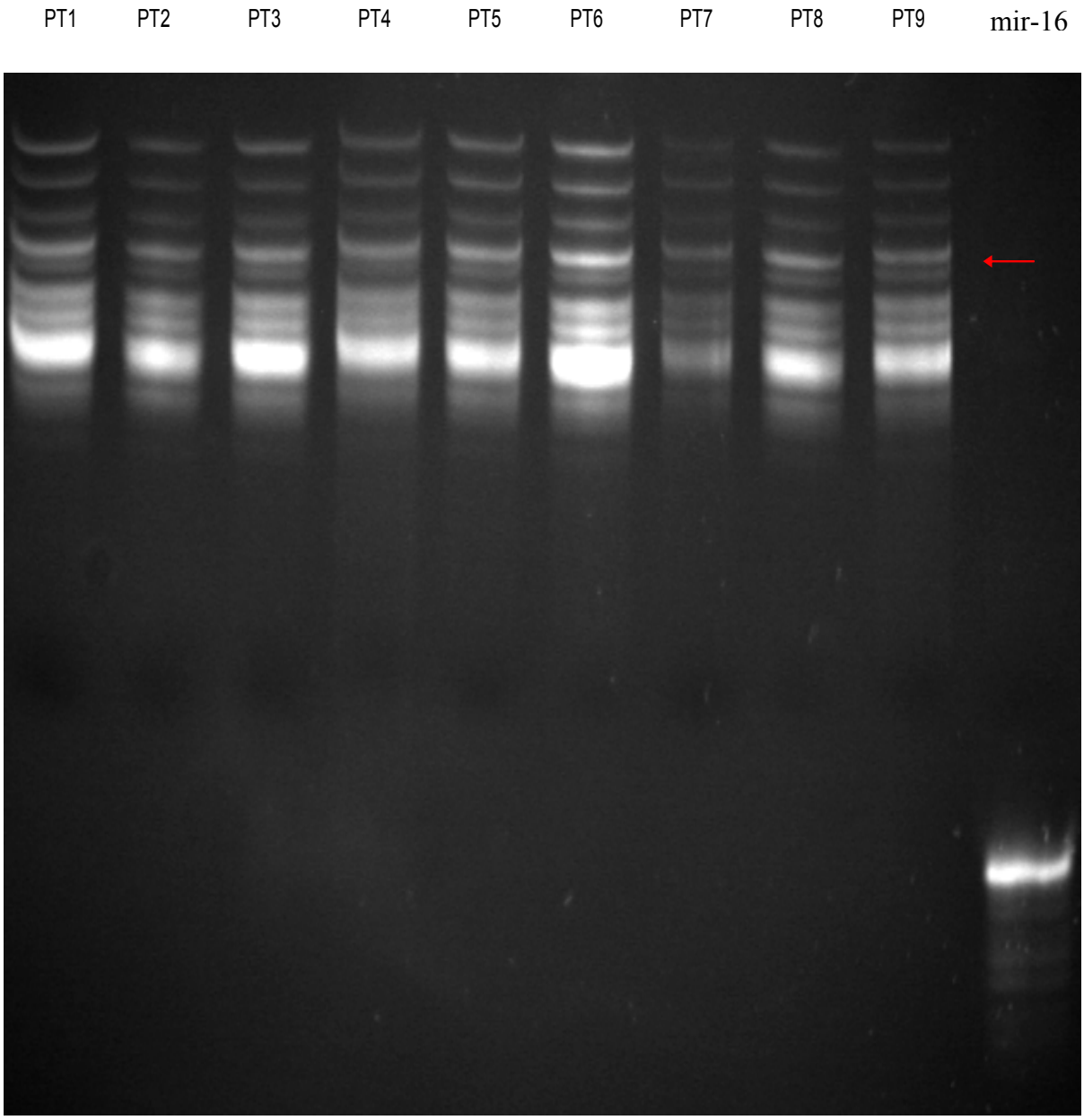

Figure 3.2. Nine miRNA samples (0.25 µg each) and standard mir-16 (0.25 µg) in the 15 % polyacrylamide gels, stained by the cybergold. The arrow shows the bands that were selected to calculate the concentration of the sample. Standard miRNA 16 which is an indicate size marker that shows the miRNA location in the gel. This standard mir-16 is shown on furthest right column.

Table 3.3. Concentration of miRNA sample of thirty patients. The twelve patient samples that are in bold font were selected for miRNA expression analysis.

PT	(ug/ul)	PT	(ug/ul)
1	**0.047**	**16**	**0.05**
2	**0.043**	**17**	**0.05**
3	**0.069**	18	0.034
4	**0.057**	19	0.036
5	**0.064**	**20**	**0.053**
6	**0.239**	21	0.032
7	0.015	22	0.018
8	**0.067**	23	0.008
9	0.026	24	0.015
10	-	25	0.014
11	0.027	26	0.018
12	**0.046**	27	0.006
13	0.032	28	0.013
14	**0.047**	29	0.016
15	0.027	30	0.015

0.5 µg of RNA is required for each solution hybridization assay. The total volume of RNA for each patient is 100 µl. In our study, we wanted to measure the expression of a minimum of eight miRNAs. Therefore, only patients that have a minimum concentration of 0.0425 µg/µl are used for miRNA expression analysis. Not all patient samples possessed RNA more than this threshold. Only twelve patients satisfied this criterion (see **Table 3.3**). Among these twelve samples, eight are in the mutated B-CLL group and four are in the unmutated B-CLL group.

3.1.3 Selection of miRNAs for Analysis

Due to the limited amount of patient RNA, we were limited to analyzing eight miRNAs in this study. To choose the miRNAs for our analysis, we focused on miRNAs that are located near genomic aberrations that are associated with B-CLL.

Döhner and Stilgenbauer reported the following B-CLL associated genomic aberrations: 55 % d13q14, 18 % d11q22-23, 16 % +12, 7 % d17p13, 7 % d6q21, t14q32, +19, where d: deletion, +: trisomy, t: translation and the percentage represents the frequency of these aberrations in their B-CLL patients (Döhner and Stilgenbauer, 2000; Stilgenbauer et al., 2003). Individual genomic aberrations relevant to mutated B-CLL or unmutated B-CLL cells have also been reported. For example, +19, d13q14 are relevant to mutated B-CLL and d11q22-23, +12, and d6q21 are relevant to unmutated B-CLL (Hamblin et al., 1999; Carsten et al., 2003).

We used the UCSC Genome database to search for miRNAs that were located at sites of genomic aberrations (http://genome.ucsc.edu/cgi-bin/hgTracks). We found 72 miRNAs that were located at specific chromosomal aberrations (Appendix F). To limit this list further, we determined miRNA target genes using Target Scan database (http://genes.mit.edu/targetscan/index.html). We identified thirty miRNAs that have target genes associated with B-CLL or other cancers (Appendix G). Finally, based on genomic location and miRNA target, we narrowed the list down to eight miRNAs (Appendix H). Specifically, we chose the following miRNAs with their associated chromosomal aberration: mir16-1 (deletion 13); let-7i, mir196-2 and mir26a-2 (trisomy 12); mir-34b and mir-125b (deletion 11); mir-181C and mir-125a (trisomy 19).

3.1.4 Solution Hybridization Assay

We used the solution hybridization assay developed by Ambion (mirVana miRNA detection kit) to monitor the expression of specific miRNAs in the patient samples. In

total, we tested the expression of eight miRNAs in eight unmutated and four mutated B-CLL patients.

Three of the eight miRNAs are not expressed in B-CLL patient samples (mir196, mir-125a, mir-125b, Data not shown). **Figures 3.3-3.7** show the solution hybridization assay for five miRNAs (mir16-1, let-7i, mir26a-2, mir-34b, and mir-181C).

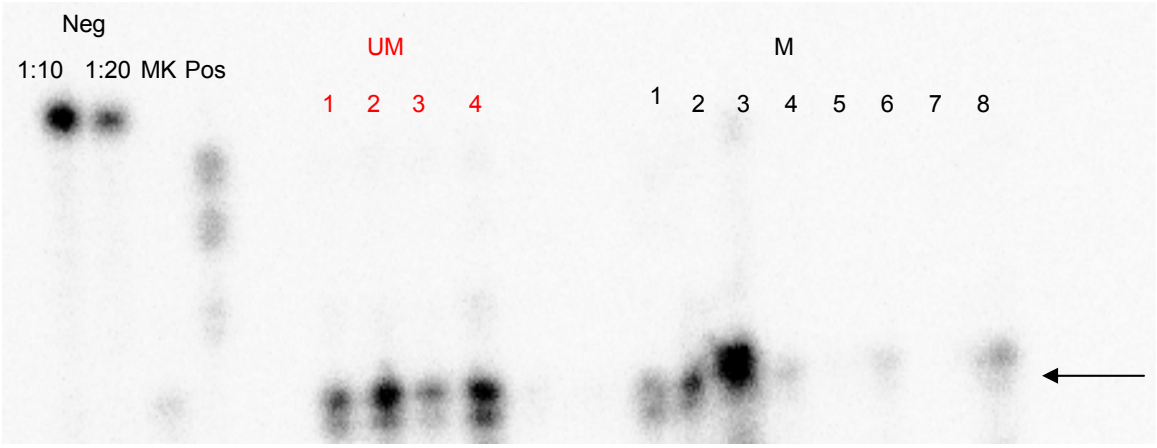

Figure 3.3. Solution Hybridization for mir-16-1. miRNA 16 solution hybridization assay products are marked with arrow. The negative controls (Neg) are undigested mir16 probes diluted at 10 and 20 times from stock solution. mir-16 probe was used as the positive marker (MK). Positive control (Pos) is mir-16 probe with RNase digestion. Four unmutated patients (UM) and eight mutated (M) patients are shown.

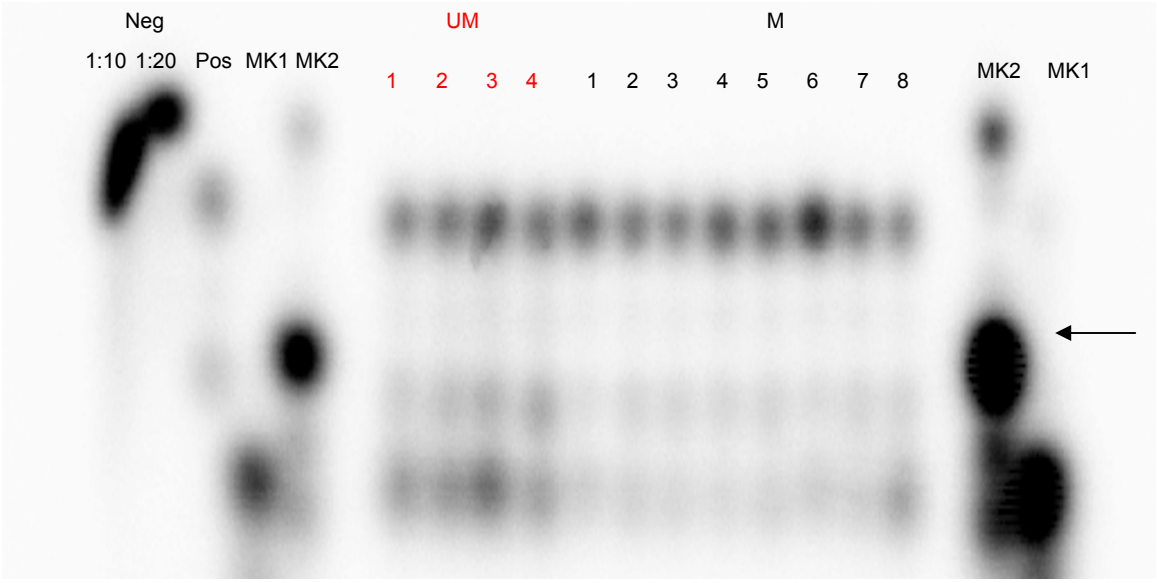

Figure 3.4. Solution Hybridization for Let-7i. Let-7i solution hybridization assay products are marked with arrow. The negative controls (Neg) are undigested mir16 probes diluted at 10 and 20 times from stock solution.mir-16 (MK1) and mir-16^{+4} (MK2) probes were used as the positive markers. Positive control (Pos) is mir-16 probe^{+4} with RNase digestion. Four unmutated patients (UM) and eight mutated (M) patients are shown.

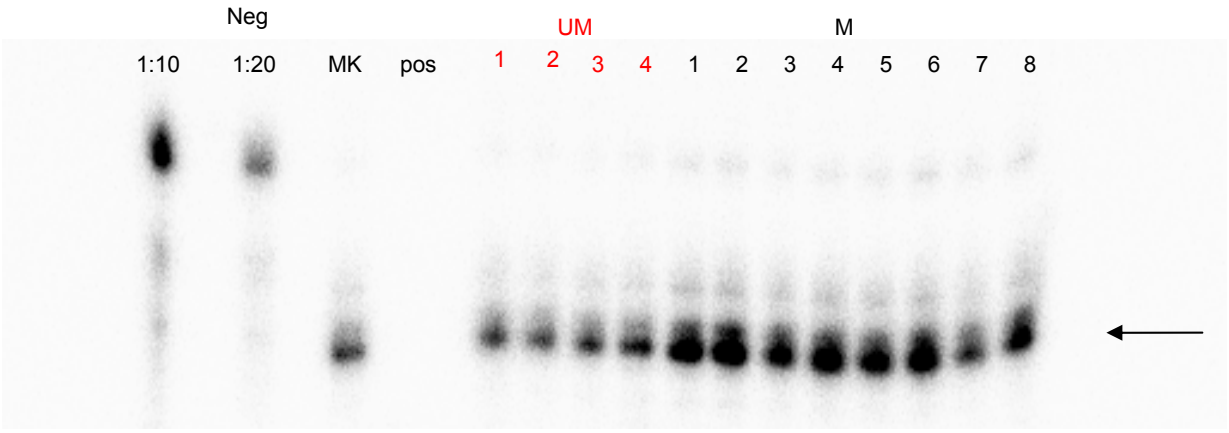

Figure 3.5. Solution Hybridization for mir-26a-2. miRNA 26a-2 solution hybridization assay products are marked with arrow. The negative controls (Neg) are undigested mir16 probes diluted at 10 and 20 times from stock solution. mir-16 probe was used as the positive marker (MK). Positive control (Pos) is mir-16 probe with RNase digestion. Four unmutated patients (UM) and eight mutated (M) patients are shown.

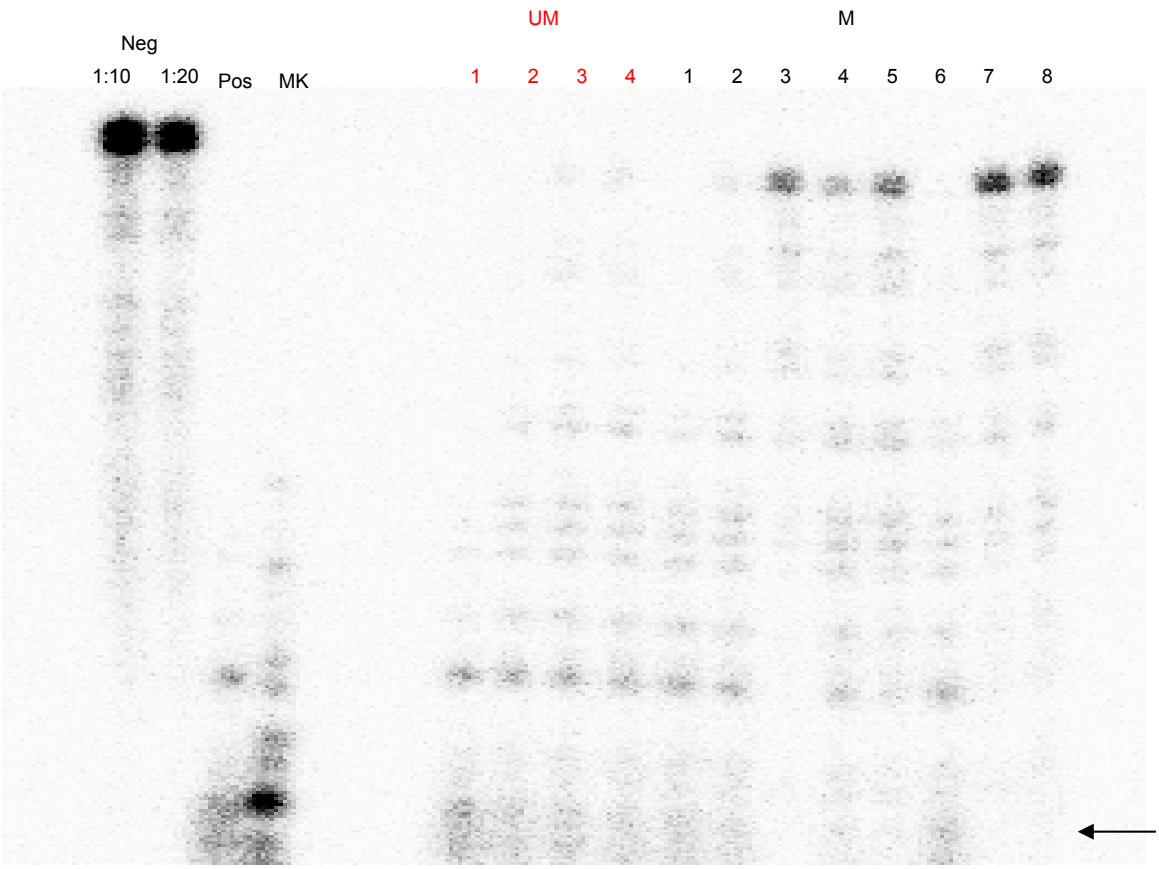

Figure 3.6. Solution Hybridization for mir-34b. miRNA 34b solution hybridization assay products are marked with arrow. The negative controls (Neg) are undigested mir-16^{+4} probe diluted at 10 and 20 times from stock solution. mir-16^{+4} probe was used as the positive marker (MK). Positive control (Pos) is mir-16^{+4} probe with RNase digestion. Four unmutated patients (UM) and eight mutated (M) patients are shown.

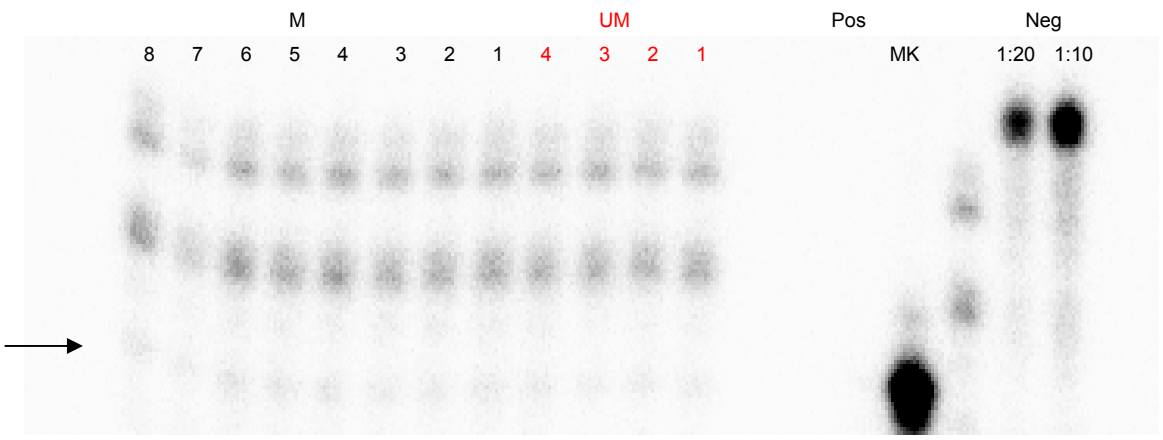

Figure 3.7. Solution Hybridization for mir-181c. miRNA 181c solution hybridization assay products are marked with arrow. The negative controls (Neg) are undigested mir-16^{+4} probe diluted at 10 and 20 times from stock solution. mir-16^{+4} probe was used as the positive marker (MK). Positive control (Pos) is mir-16^{+4} probe with RNase digestion. Four unmutated patients (UM) and eight mutated (M) patients are shown.

3.1.5 SAS Analysis

We analyzed the expression of eight miRNAs in twelve patient samples. We measured the intensity of the miRNA bands using a phosphoroimager (Bio Rad Inc). We used t-statistic analysis to determine a P-value describing the probability that mutated and unmutated B-CLL groups differentially expresses a chosen miRNA. We defined miRNA as differentially expressed in the two groups if the P-value is less than 0.05. The complete SAS analysis is in Appendix I. The results are summarized in **Table 3.4.** The following miRNAs (**mir16-1, mir26a-2, let-7i, mir-34b**) have a P-value less than 0.05 and therefore they are differentially expressed in the mutated and unmutated B-CLL groups. **mir-181c** has a P-value of 0.08 and therefore it is not differentially expressed in the mutated and unmutated B-CLL groups. In summary, determined that mir26a-2 expression is up-regulated in the mutated B-CLL group and **mir16-1, let-7i, mir-34b** are down-regulated in the mutated B-CLL group.

Table 3.4. Eight selected miRNA expressions

miRNA	M-CLL (mean)	UM-CLL (mean)	Location	Type*	P
mir16-1	76468	194760	13q14	Down	0.0489
mir196-2	n/a	n/a	12q13	n/a**	n/a
mir26a-2	3.41E+06	1.74E+06	+12q14.1	Up	0.0107
let-7i	706010	1.29E+06	12q14.2	Down	0.0296
mir-34b	76963	166256	11q23	Down	0.0265
mir-125b	n/a	n/a	11q24.1	n/a	n/a
miR-181c	338837	268287	19p13.3	ND***	0.0802
miR-125a	n/a	n/a	19q13.4	n/a	n/a

*: Take UM-CLL (unmutated B-CLL) as a reference; ** n/a: not available; ***ND: no difference.

3.2 Discussion

Our study aimed to discover miRNAs that classify the mutated and unmutated sub-groups of B-CLL. We chose miRNAs since they are involved in developmental process (Baehrecke, 2003; Bartel, 2004; Calin et al., 2004) and are associated with specific cancers (Chena and Lodish, 2005; Michael et al., 2003). Many miRNAs are also located at sites of B-CLL chromosomal abnormalities (Calin et al., 2002). B-CLL chromosomal abnormalities with related miRNAs analyzed in this study include: mir16-1 (deletion 13); let-7i, mir196-2 and mir26a-2 (trisomy 12); mir-34b and mir-125b (deletion 11); mir-181C and mir-125a (trisomy 19).

Our experimental results show that the following miRNAs (mir16-1, mir26a-2, let-7i, mir-34b) have significant differences in their expressions between the mutated and unmutated B-CLL groups. During the completion of this study, Calin et al., (2004) used

microarrays to perform genome wide expression profiling of miRNAs. They identified the following miRNAs that distinguished mutated and unmutated B-CLL: mir-186, mir-132, mir-16-1, mir-102, and mir-29c. In Calin et al., (2004) and our study, mir-16-1 is the only common miRNA that distinguishes mutated and unmutated B-CLL.

CHAPTER 4

Conclusions and Future Work

This thesis presents a study on miRNA expression in two sub-groups of B-CLL: mutated and unmutated B-CLL. The main purpose of the study is to examine whether miRNAs can be used as surrogate markers for the unmutated IgV_H group of B-CLL patients. We analyzed a specific set of miRNAs that were selected based on their chromosome location and mRNA targets. In total, we analyzed the following eight miRNAs, which are shown together with their associated chromosomal abnormalities: deletion 13: mir16-1; trisomy 12: let-7i, mir196-2, mir26a-2; deletion 11: mir-34b, mir-125b; trisomy 19: mir-181C, mir-125a. We determined the mutational status of thirty patient samples and analyzed miRNA expression in twelve of these patients (eight M-CLL and four M-CLL). The following conclusions can be drawn from our experiment:

(1) miRNAs **mir16-1, mir26a-2, let-7i, and mir-34b** have significant differences in expressions in mutated and unmutated B-CLL subgroups. The confidence level in our statistical inference is 95 %.

(2) miRNA **mir-181c** has no significant difference in its expression over the two B-

CLL group.

(3) miRNA **mir-16-1** has also been reported to be differentially expressed in mutated and unmutated B-CLL (Calin et al., 2004). Therefore, **mir-16-1** is a reliable surrogate marker for IgV_H mutational status.

The present work shows promises for finding differential miRNA expression in subtypes of B-CLL. This can not only be used as an alternative way to sub-classify the disease, but also have the potential to discover the molecular mechanisms of the disease. Some future work may be suggested. (i) Development of a classifier based on miRNAs, as a surrogate marker for the unmutated IgV_H gene, may be needed in order to increase the reliability of this new marker. Nevertheless, it is likely that the miRNA, mir-16-1, alone could be an excellent marker. (ii) Global miRNA profiling using microarray will be used to screen large sets of human miRNAs in B-CLL. (iii) Correlation of the miRNAs expression with their chromosomal abnormalities to determine whether specific translocations up or down-regulate miRNA expression.

REFERENCE

Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T. and Jewell, D. (2003). MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Curr. Biol.* **13**, 807-818.

Andrea, T. and Peter, P.T. (2004). Molecular evolution of a microRNA cluster. *J. Mol. Biol.* **339**, 327–335.

Aravin, AA., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J. and Tuschl, T. (2003). The small RNA profile during drosophila melanogaster development. *Dev Cell.* **5**, 337-350.

Ausubel FM., Brent, R., Kingston, RE. and Moore, DD. (1994-2004). Current Protocols in Molecular Biology. vol.1, *John Wiley and Sons, Inc.*, *Brooklyn*. New York. 3.13.1-3.13.3.

Baehrecke EH. (2003). miRNAs: micro managers of programmed cell death. *Curr Biol,* **13,** R473-475.

Bartel DP. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. **116**, 281-297.

Berk K.N and P Carey. (2004). Data Analysis with Microsoft Excel. *Duxbury, Pacific Grove*. *CA*. 579pp.

Binet JL., Auquier, A., Dighiero, G., Chastang, C., Piguet, H., Goasguen, J., Vaugier, G., Potron, G., Colona, P., Oberling, F., Thomas, M., Tchernia, G., Jacquillat, C., Boivin, P., Lesty, C., Duault, MT., Monconduit, M., Belabbes, S. and Gremy, F. (1981). A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. *Cancer*. **48,** 198-206.

Brian T., Pittner, RC., Tschumper, TK., Kimlinger, CM., Thomas, PJ., Kurtin, NE. and Kay, DF. (2003). Expression of ZAP-70 in Both Unmutated and Mutated Leukemic CLL B Cells Indicates Lack of Efficacy as a Surrogate Marker for Immunoglobulin Mutational Status. *ASH,* **109** and *Blood*, **102,** 11.

Carsten S., Michelle, N. and Swen, W. (2003). Automated array-based genomic profiling in Chronic Lymphocytic Leukemia, Development of a clinical tool and discovery of recurrent genomic alterations. *Proc Natl Acad Sci.* USA, **101**, 1039-1044.

Calin GA., Sevignani, C. and Dumitru, CD. (2004). Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci*. **101**, 2999-3004.

Calin GA., Dumitru, D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating M. and Rai, K. (2002). Frequent deletions and down-regulation of miRNA

gene mir-15 and mir16 at 13q14 in chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. USA*. **99,** 15524–15529.

Chena, C-Z. and Lodish, H.F. (2005). MicroRNAs as regulators of mammalian hematopoiesis. *Seminars in Immunology*. **17**, 155–165.

Cheson, BD., Bennett, JM., Grever, M., Kay, N., Keating, MJ., O'Brien. S. and Rai, KR. (1996). National Cancer Institute-sponsored Working Group guidelines for chronic lymphocytic leukemia, revised guidelines for diagnosis and treatment. *Blood.* **87**, 4990-4997.

Chirgwin, J.M., Przybyla, A.E., MacDonald R.J. and Rutter, W.J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry*. **18,** 5294-5299.

Damle, T., Wasil, T., Fais, F., Ghiotto, F., Valetto, A., Allen, SL., Buchbinder, A., Budman, D., Dittmar, K., Kolitz, J., Lichtman, SM., Schulman, P., Vinciguerra, VP., Rai, KR., Ferrarini, M. and Chiorazzi, N. (1999). Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood*. **94,** 1840-1847.

Devore, J. and Nicholas, F. (1999). Applied statistics for engineers and scientists. *DUXBURY PRESS. USA*.

Döhner, H., Stilgenbauer, S., Benner, A., Leupolt, E., Krober, A., Bullinger, L., Dohner, K., Bentz, M. and Lichter, P. (2000). Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med.* **343**, 1910-1916.

Dostie, J., Mourelatos, Z., Yang, M., Sharma, A. and Dreyfuss, G. (2003). Numerous microRNPs in neuronal cells containing novel microRNAs. *RNA*. **9**, 631-632.

Durig, J., Nuckel, H. and Cremer, M. (2003). ZAP-70 expression is a prognostic factor in chronic lymphocytic leukemia. *Leukemia*. **17,** 2426-2434.

Fais, F., Ghiotto, F., Hashimoto, S., Sellars, B., Valetto, A., Allen, SL., Schulman, P., Vinciguerra, VP., Rai, K., Rassenti, LZ., Kipps, TJ., Dighiero, G., Schroeder, HW., Ferrarini Jr.M. and Chiorazzi, N. (1998). Chronic lymphocytic leukemia B cells express restricted set of mutated and unmutated antigen receptors. *J Clin Invest*. **102**, 1515-1525.

Gaiger, A., Heintel, D. and Jäger.Novel, U. (2004). August. molecular diagnostic and therapeutic targets in chronic lymphocytic leukaemia. *European Journal of Clinical Investigation.* **34**, no. s2, 25-30.

Gozzetti, A., Crupi, R. and Tozzuoli, D. (2004). The use of fluorescence in situ hybridization (FISH) in chronic lymphocytic leukemia (CLL). *Hematology*. **9**, 11-15.

Hallek, M., Langenmayer, I., Nerl, C., Knauf, W., Dietzfelbinger, H., Adorf, D., Ostwald, M., Busch, R., Kuhn-Hallek, I., Thiel, E. and Emmerich B. (1999). Elevated serum thymidine kinase levels identify a subgroup at high risk of disease progression in early, nonsmoldering chronic lymphocytic leukemia. *Blood*. **93**, 1732.

Hamblin, TJ., Davis, Z., Gardiner, A., Oscier, DG. and Stevenson, FK. (1999). Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood*. **94**, 1848-1854.

Han, T., Bhargava, A. and Henderson, ES. (1989). Prognostic significance of beta-2-microglobulin (β-2m) in chronic lymphocytic leukemia (CLL) and non-Hodgkin's lymphoma (NHL). *Proc Am Soc Onco*. **l8,** 270.

Ivan, A. and Ian, W.F. (2003). Chronic lymphocytic leukemia: advances in biology and therapeutics. *Current Opinion in Oncology*. **15**, 16-22.

Kim, J., Krichevsky, A., Grad, Y. and Hayes, GD., (2003). Identification of many microRNAs that copurify with polyribosomes in mammalian neurons. *Proc. Natl. Acad Sci, USA. Proc Natl Acad Sci*. **10**, 1073.

Kröber, A., Seiler, T., Leupolt, E., Dohner, H. and Stilgenbauer, S. (2000). IgV_H mutated and unmutated B-CLL tumors show distinct genetic aberration patterns. *Blood.* **96** (Suppl 1), 835a.

Küppers, R., Klein, U., Hansmann, ML. and Rajewsky, K. (1999). Cellular origin of human B-cell lymphomas. *N Engl J Med*. **341**, 1520-1529.

Lagneaux, L., Delforge, A., Bron, D., Bruyn, C.D. and Stryckmans, P. (1998). Chronic Lymphocytic Leukemic B Cells But Not Normal B Cells Are Rescued From Apoptosis by Contact With Normal Bone Marrow Stromal Cells. *Blood.* **91,** 2387-2396.

Lee, Y., Jeon, K., Lee, JT., Kim, S. and Kim, VN. (2002). MicroRNA maturation, stepwise processing and subcellular localization. *EMBO J.* **21**, 4663–4670.

Lossos, IS., Tibshirani, R., Narasimhan, B. and Levy, R. (2000). The inference of antigen selection on Ig genes. *J Immunol*. **165**, 5122-5126.

Lund, E., Guttinger, S., Calado, A., Dahlberg, JE. and Kutay, U. (2004). Nuclear export of microRNA precursors. *Science.* **303**, 95–98.

Matutes, E., and Catovsky, D., (1993). CLL should be used only for the disease with B-cell phenotype. *Leukemia,* **7**, 917-918.

Michael, Z., Connor, M., Young P. and James, J. (2003). Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Mol. Cancer Res.* **1,** 882–891.

Moss E.G. (2003). MicroRNAs in Noncoding RNAs. *Molecular Biology and Molecular Medicine.* **17,** 98–114.

Mourelates, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M. and Dreyfuss, G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.* **16**, 720-728.

Nelson, P., Marianthi, K., Anup, S., Elsa, M. and Zissimos, M. (2003). The microRNA world: small is mighty. *Trends Biochem Sci*. **28**, 534–540.

Rai, KR., Sawitsky, A. and Cronkite, EP. (1975). Clinical staging of chronic lymphocytic leukemia. *Blood*. **46,** 219-234.

Rosenwald, A., Ash, A., Simon, R., Davis, R.E., Yu, X., Yang, L., Powell, J., Botstein, D., Michael, R.G., Bruce, D.C., Nicholas, C., Wyndham, H.W., Thomas, J.K., Patrick. O.B. and M Staudt, L. (2001). Relation of Gene Expression Phenotype to Immunoglobulin Mutation Genotype in B Cell Chronic Lymphocytic Leukemia. *The Journal of Experimental Medicine* . **194**, 1639–1647.

Rozman, C. and Montserrat, E. (1995). Chronic lymphocytic leukemia. *New England Journal of Medicine*. **333**, 1052–1057.

Schroeder, HW.Jr., and Dighiero, G. (1994). The pathogenesis of chronic lymphocytic leukemia: analysis of the antibody repertoire. *Immunol Today*. **15,** 288-294.

Schwarz, D.S. and Zamore, P.D. 2002. Why do miRNAs live in the miRNP. *Genes Dev*. **16**, 1025–1031.

Stilgenbauer, S., Döhner, H. and Lichter, P. (2003). Evaluation of the Clinical Impact of Expression Profiling in Chronic Lymphocytic Leukemia. *Molecular Medicine.* **15,** 156-157.

Stilgenbauer, S., Lichter, P., and Dohner, H .2000. Genetic features of B-cell chronic lymphocytic leukemia. *Rev Clin Exp Hematol*. **4**, 48-72.

Thunberg, U., Johnson, A. and Roos, G. (2001). CD38 expression is a poor predictor for VH gene mutational status and prognosis in chronic lymphocytic leukemia. *Blood*. **97**, 1892-1893.

Appendix A: The significance test and SAS program

A.1 Theory

One of the statistical analyses is to compare two groups, e.g., mutated B-CLL (M-CLL) and unmutated B-CLL (UM-CLL). Physical, chemical or biological properties or behaviors of the two groups are called responses or variables. Each group has a number of elements (or samples in the terminology of applications). Each element gets a response. The average of the responses of all elements in a group is one of the statistics of the group, which is called mean. The difference between two groups can then be represented by the difference between two means. Note that the response or variable is random, i.e., what value of the variable for a particular element is uncertain. As such, the mean is also a random variable in nature. This implies that the means of two groups will always be of different values. In other words, we cannot get any conclusion about whether two groups are different or not based on an arithmetic evaluation of the difference of their means (they are always different!). Nevertheless, the statistics concerns what is so-called "significant" difference. The test for significant difference between two groups is generally called the significance test.

Suppose we have two groups (1, 2). We have n_1 elements for group 1 and n_2 elements for group 2. Then we have a mean for group 1 (μ_1) and a mean for group 2 (μ_2). The significance test is based on the following procedure.

There are two types of hypotheses defined as follows.

(i) The null (H_0): $\mu_1 = \mu_2$

(ii) The alternative (H_a): $\mu_1 \neq \mu_2$

H_0 states that there is no significant difference between two groups, while H_a states that there is significant difference between two groups.

To test the hypotheses above, we define statistics quantities (or statistics if confusion does not arise). The most popular statistics in this case is the t-statistics. The t-test is named for the testing of the aforementioned hypotheses using the t-statistics. The t-statistics is defined as follows:

(i) The variances of the two groups are equal (Pooled model):

$$DF = n_1 + n_2 - 2 \tag{I-1}$$

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})}{n-1} \qquad (S=S_1 \text{ when } n=n_1;\ S=S_2 \text{ when } n=n_2) \tag{I-2}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)} \tag{I-3}$$

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}} \tag{I-4}$$

In the above S is the standard deviation of the sample, X is the response of an element in a group, and $\overline{X}$ is the mean of the responses of all elements in the group. It is noted that in statistics, we have the concept: population and sample. Both are, mathematically, a set which contains elements. Yet, the sample is a portion of the population, as we may not be able to get all elements of the population. Usually, the mean of a population is denoted by μ; while the mean of a sample is denoted by X.

(ii) The variances of the two groups are unequal (Satterthwaite Model):

$$DF = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2 - 1}} \qquad \text{(I-5)}$$

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})}} \qquad \text{(I-6)}$$

From the above equations, in particular with the t-value and DF value, we can determine the P-value from the standard table that can be found from any statistics text book (e.g., Berk and Carey, 2004). If a computer aided statistic software program (e.g., SAS) is used, the P-value can be calculated by that program automatically. The P-value is the probability of the event that the difference between two means is due to some chance-based causes rather than some structural-based causes. Therefore, the smaller

the P-value the more likelihood the H_0 hypothesis is rejected (i.e., in other words two groups have significant difference based on their means). There are a couple of things to be remarked.

First, in the above procedure, the meaning of "small" needs to be defined. This is done by giving the so-called confidence level or significant level; often this value takes 0.10, 0.05, 0.01 depending on the application domain. The significance level has the same value of the type I error in the hypothesis test above, denoted by α. The type I error is an error that H_0 is correct but is wrongly rejected.

Second, in the above procedure, the procedure to determine the equal variance or unequal variance is the Levene test procedure. The Levene test is defined as:

H_0: $\sigma_1 = \sigma_2 = \ldots = \sigma_k$

H_α: $\sigma_i \neq \sigma_j$ for at least one pair (i. j).

where σ_i is the standard deviation of sample i (or group i), and k is the total number of groups. The test statistics used for the Levene test is the F-statistics which is defined as

$$\text{F-statistics} = \frac{BSV}{WSV} \qquad \text{(I-7)}$$

where BSV: Between-sample-variation, and

WSV: Within-sample-variation.

The Levene test then proceeds as follows. Given samples or groups (1, 2, …, k) we can calculate the F-statistics, say F=1.7059. After that, we have two alternative ways to test

the hypothesis. One way is to find the critical F-value (denoted by Fc) given the significant level (e.g. α=0.05). The other way is to find the P-value. Note that the standard programs are available to perform both calculations. By taking the first way, if F > Fc we reject the null hypothesis at α level (e.g. α=0.05) (i.e., the unequal variance situation). By taking the second way, if P-value < α, we reject the null hypothesis.

Third, there are two assumptions behind the t-test: (i) the dependent variable (e.g., the miRNA expression) should be continuous and follow the normal distribution, and (ii) the independent variable (e.g., patient sample) should be mutual exclusive (i.e., each element appears either in group 1 or group 2, but not both). For a particular t-test program, any departure away from the assumptions may lead to error in the test.

Last, the meaning of difference of two quantities (e.g., μ_1 and μ_2) could be: (i) $\mu_1 \neq \mu_2$, (ii) $\mu_1 > \mu_2$, and (iii) $\mu_1 < \mu_2$. The (i) is called two-tail, while (ii) and (iii) are called one-tail. As a convention, the standard program will take the two-tail as a default when it comes out with the P-value and F-statistics. If a particular problem is on-tail, the corresponding P-value and F-statistics must divide by 2.

A.2 Software

SAS is a program which implements many statistical analysis methods, including the significance test. In SAS, a particular module or procedure called ‘t-test’ can be used to perform the significance test based on the t-statistics (t-test for short). The following

example illustrates the input and output data files of SAS for the t-test which includes the Levene test.

Table A.1 shows an example of miRNA (mir-26a). The SAS input for this example is given in **Figure A.1**. The SAS output file is given in **Figure A.2**. From the lower portion of **Figure A.2** we can find (i) F = 9.42, and (ii) {Pr >F} = 0.0924. Now because {Pr >F}> 0.05, the null hypothesis is accepted; in other words, the two groups in the example have any equal variance. Then we read out from the output T-Tests, specifically the line for the equal variance, the following: {Pr > |t|} = 0.0107. Since {Pr > |t|} < 0.05, we conclude that the two groups have significance difference in their means. It is noted that the terms ('std Dev', 'std Err', 'DF', 'Num DF', 'Den DF') are not informative to the purpose of the t-test.

Table A.1. The expressions of mir-26a in the two groups (UM, M)

Rep	M	UM
1	4271839.8	
2	5154553.9	
3	2639191	
4	3606797	
5	3216792.2	
6	3688059.3	
7	1894745.6	
8	2793187.3	
9		1421870.9
10		1568594.1
11		1803838.1
12		2179905.6

```
data first;
input Rep species$ intensity;
Cards;
1       M       4271839.8
1       M       5154553.9
1       M       2639191
1       M       3606797
1       M       3216792.2
1       M       3688059.3
1       M       1894745.6
1       M       2793187.3
2       UM      1421870.9
2       UM      1568594.1
2       UM      1803838.1
2       UM      2179905.6
;
proc ttest;
      class species;
      var   intensity;
            title 'T-test using SAS ttest';
run;
```

Figure A.1. The SAS input file for mir-26a expression

T-test using SAS ttest 10:56 Friday, August 13, 2004 1
The TTEST Procedure
Statistics

Variable	patients	N	Mean	Mean	Mean	Std Dev	Std Dev	Std Dev	Std Err
intensity	M	8	2.56E6	3.41E6	4.26E6	671027	1.01E6	2.07E6	358822
intensity	UM	4	1.22E6	1.74E6	2.27E6	187350	330720	1.23E6	165360
intensity	Diff (1-2)		479930	1.66E6	2.85E6	606650	868235	1.52E6	531683

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
intensity	Pooled	**Equal**	10	3.13	**0.0107**
intensity	Satterthwaite	Unequal	9.31	4.21	0.0021

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
intensity	Folded F	7	3	9.42	0.0924

Figure A.2. The SAS output file for mir-26a expression

Appendix B: RT-PCR results

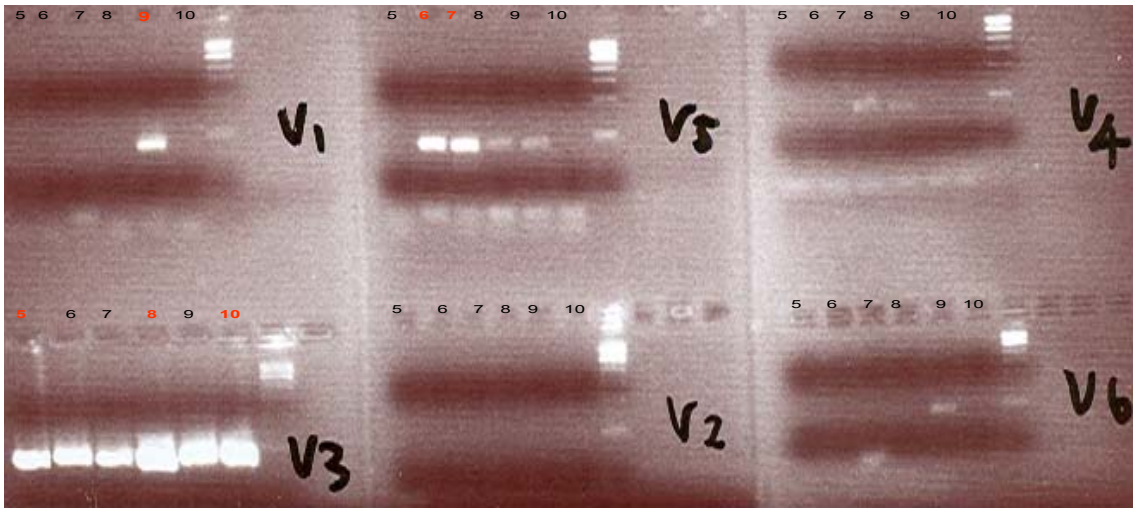

Figure B.1. The RT-PCR products from PT5-PT10 analyzed on a 2 % agarose gel, stained with the ethidium bromide.

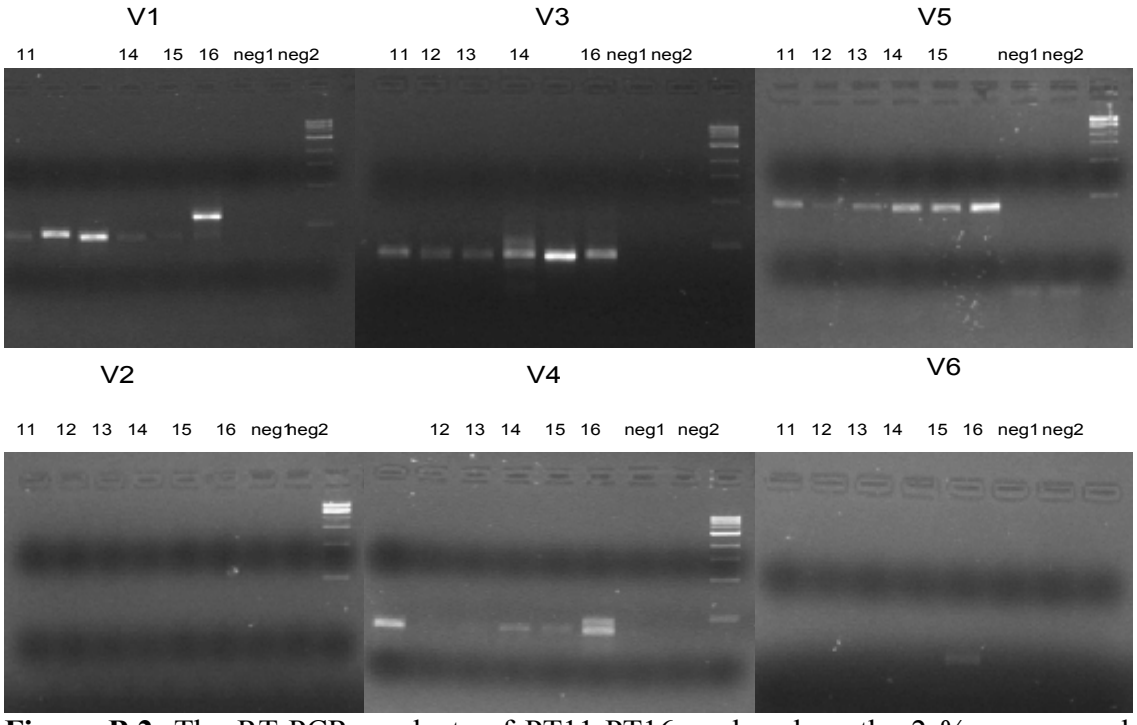

Figure B.2. The RT-PCR products of PT11-PT16 analyzed on the 2 % agarose gel, stained with the ethidium bromide.

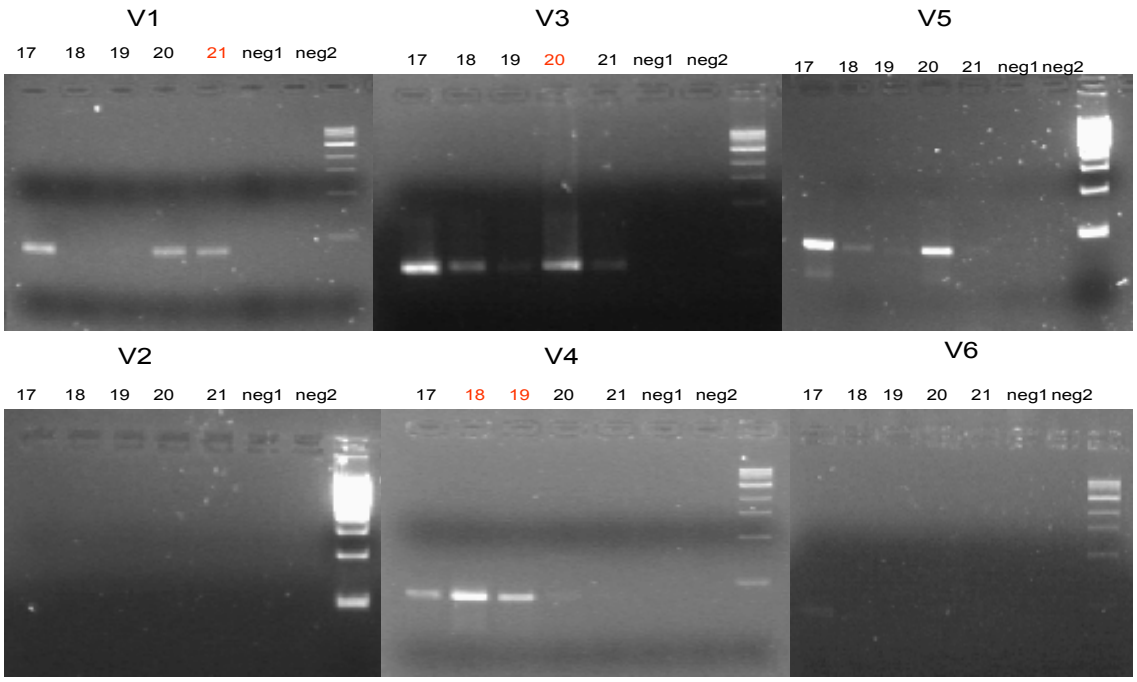

Figure B.3. The RT-PCR products of PT17-PT21 analyzed on 2 % agarose gel, stained with the ethidium bromide.

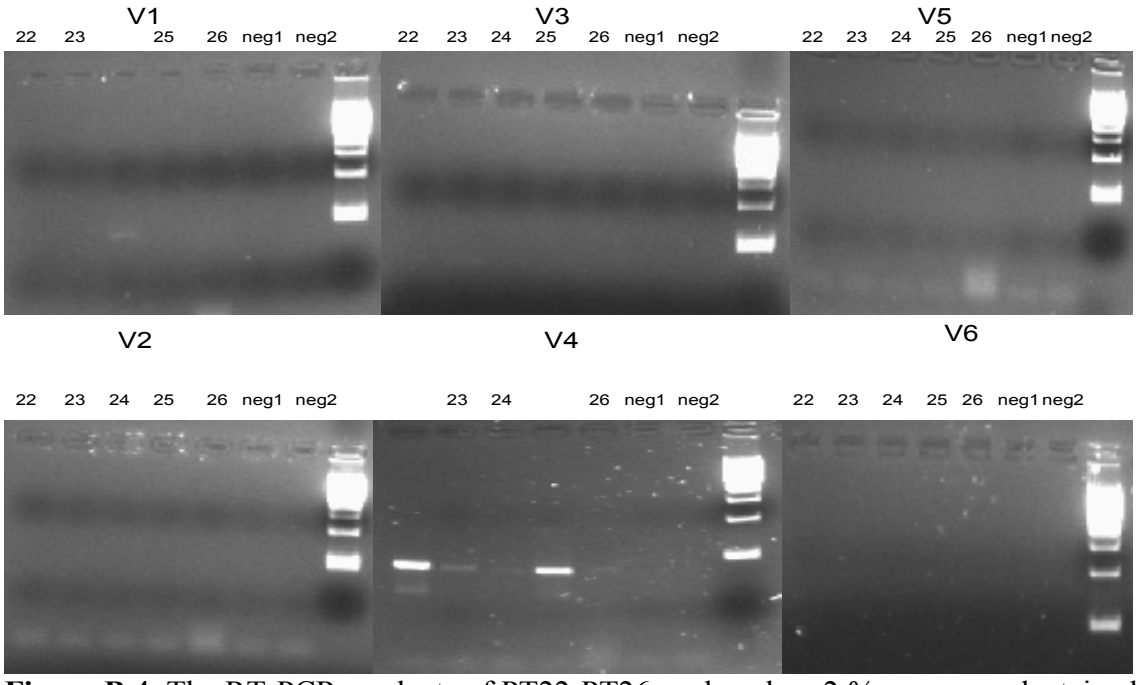

Figure B.4. The RT-PCR products of PT22-PT26 analyzed on 2 % agarose gel, stained with the ethidium bromide.

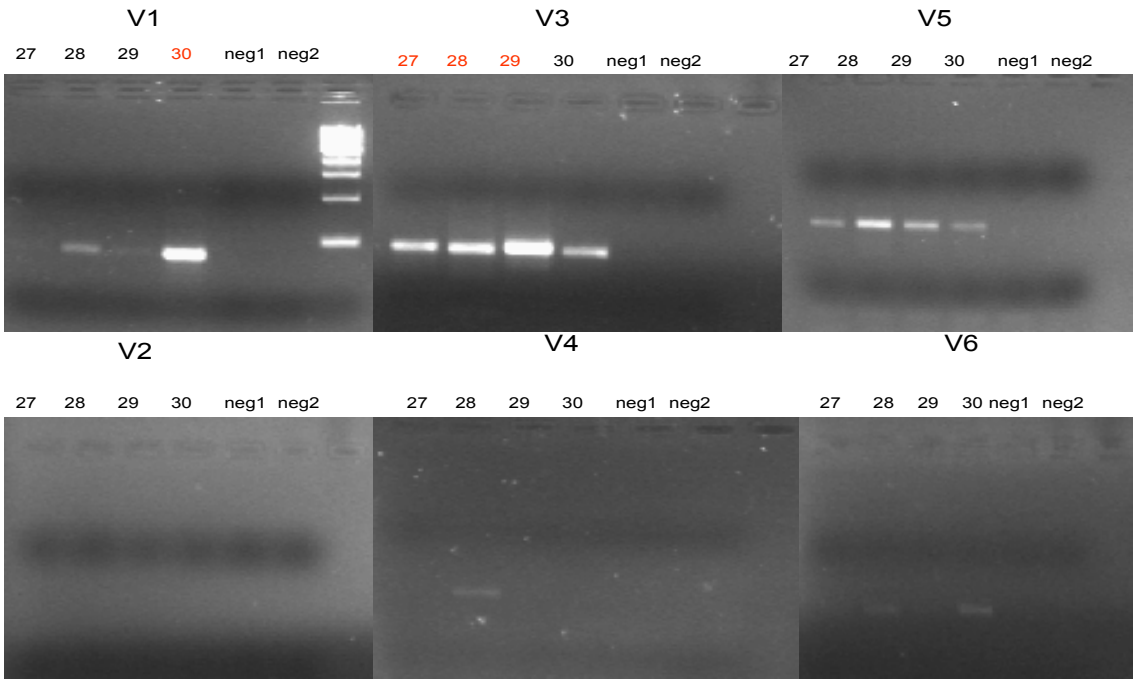

Figure B.5. The RT-PCR products of PT27-PT30 analyzed on 2 % agarose gel, stained with the ethidium bromide.

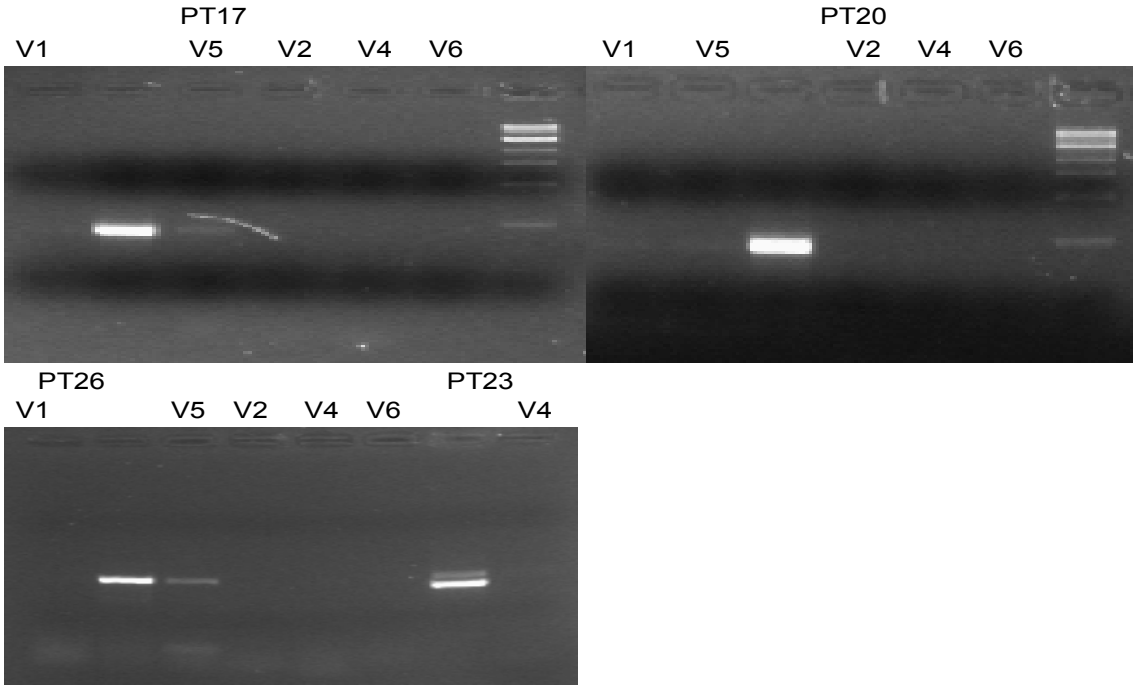

Figure B.6. The reanalysis of RT-PCR products for PT17, PT20, PT 23, and PT26 on a 2 % agarose gel, stained with the ethidium bromide.

Appendix C: The IGV_H gene sequences of 30 Patients

PT1

CCAGATGGGTCTGTCCAGGTGCAGCTCAGCAGTGGGGCNCAGGACTGTTGA
AGCCTTCGGAGACCCTGTCCCTCACCTGCGCTGTCTATGGTGGGTCCTTCAG
TGATTACTACTGGACCTGGGTCCGCCAGCCCCCAGGGGGGGGTCTGGAGTG
GATTGGGNAAATCCATCACATTGGAGGTGCCAAATACAATCCGTCCCTCAA
GAGTCCAGTCACCATATCAATGGACACGTCCAAGAGCCAGTTCTCCCTGAG
ACTGATTTCTGTGACCGCCGCGGACACGGCTGTATATTACTGTGCGAGACTC
CCTCCGGAGTGGTTATTTCCTTTTGACTCCTGGGGCCCGGGGAACCCCGGTC
ACCGTCTCCTCAGGTAC

PT2

CCTCCTCCTGGCTGTTCTCCAAGGAGTCTGTGCCGAGGTGGCAGCTGGTGCA
GTCTGGAGCAGAGGTGAAAAAGCCCGGGGAGTCTCTGAAGGTCTCCTGTCA
GGGTTCTGGATTCAGCTTTACCAGCTATTGGATCGGCTGGGTGCGCCAGATG
CCCGGGAAAGGCCTGGAGTGGATGGGAATCATCGATCCTGATGACTCTGAT
ACCAGATACAGCCCGTCCTTCCAAGGCCAGGTCACCATCTCAGCCGACAAG
TCCATCAGCACCGTCTACCTGCAGTGGGGCAACCTGAGGGCCTCGGACACC
GCCATTTATTACTGTGCGAGACTCGGATATAGTTATGGGCAACTTACGGCGT
ACATCCAGCACTGGGGCCAGGGCACCCTGGTCACCGTCTCCTCAGGTAA

PT3

GGTTTTCCTTGTTGCTATTTTAAAAGGTGTCCAGTGTGAGGTGCAGCTGGTG

GAGTCCGGGGGAGGCTTAGTTCAGCCTGGGGGGTCCCTGAGACTCTCCTGT

GCAGCCTCTGGATTCACCTTCAGTAGCTACTGGATGCACTGGGTCCGCCAAC

CTCCAGGGAAGGGGCTGGTGTGGGTCTCACGTATTAATAGTGATGGGAGTA

GCACAAGGTACGCGGACTCCGTGAAGGGCCGATTCACCATCTCCAGAGACA

ACGCCAAGAACACGCTGTATCTGCAAATGAACAGTCTGAGAGCCGAGGACA

CGGCTGTGTATTACTGTGCAAGAGCCGTAGCGGGTTCGGGGAATTATTATA

ACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGTA

PT4

GGTTTTCCTCGTTGCTCTTTTAAGAGGTGTCCAGTGTCAGGTGCAGCTGGTG

GAGTCTGGGGGAGGCGTGGTCCAGCCTGGGAGGTCCCTGAGACTCTCCTGT

GCAGCGTCTGGATTCACCTTCAGTAGCTATGGCATGCACTGGGTCCGCCAGG

CTCCAGGCAAGGGGCTGGAGTGGGTGGCAGTTATATGGTATGATGGAAGTA

ATAAATACTATGCAGACTCCGTGAAGGGCCGATTCACCATCTCCAGAGACA

ATTCCAAGAACACGCTGTATCTGCAAATGAACAGCCTGAGAGCCGAGGACA

CGGCTGTGTATTACTGTGCGAGAGAATTCCTGAGGTATTACGATTTTTGGAG

TGGCTATAACGATGCTTTTGATATCTGGGGCCAAGGGACAATGGTCACCGTC

TCCNTCAGGTAA

PT5

TGGCTTTTTCTTGTGGCTTTTTAAAAGGTGTCCAGTGTGAGGTGCAGTTGTTG

GAATCTGGGGGAGGCTTGGTACAGCCGGGGGGGTCCCTGAGACTCTCCTGT

GCAGCCTCTGGATTCACCTTTAGCAGCCGTGCCATGAGCTGGGTCCGCCAGG

CTCCAGGGAAGGGGCTGGAGTGGGTCTCAGGTATTAGTGGTAGTGGAGGGA
GGACATATCACGGAGACTCCGTGAAGGGCCGGTTCACCATCTCCAGAGACA
ATTTCAAGAACACAGTGTATCTGCAAATGAACAGCCTGAGGGCCGAGGACT
CGGCCGTATATTACTGTGCGAAAGATATGTATTATAACCCTATGGACTATGA
TACCAGTGCATTTGACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCA
GGTA

PT6

AGGTTNCCTGTCTCGTNTGCAGGAGTCTGTGCCGAAGTGCAGCTGGTGCAGT
CCGGANNAGAGGTGAAAAAGCCCGGGGAGTCTCTGACGATCTCCTGTNAGG
GTTNTGGATACAGTCTTTACCAACTACTTGGATCATGCTGGGTGCGCCANAT
GCCCGGGAAAGGCCTGGAGTGGATGGGNAGNATTGATCCTAGTGACTCTTA
TACCAANTACAGCCCGTCCTTCCAAGGCCNCGTCANCATCTCAGTTGACAA
GTCCTATCATCACTGCCTACCTGCAGTGGAGCANCCTGAAGGCCTCGGACA
CTCGCCATGTATTACTGTGCGAGACTTCTCAGGNANGGACNACTGGTTGCCA
NGCT

PT7

CAAGGAGTCTGTGCCGAGGTGCAGCTGGTGCAGTCTGGAGCAGAGGTGAAA
AAGCCCGGGGAGTCTCTGAAGATCTCCTGTAAGGGTTCTGGATACAGCTTTA
CCAANTACTGGATCAGGCTGGGTGCGCCAGATGCCCGGGAAAGGCCTGGAG
TGGATGGGGATCATNTATCCTGGTGACTCTGATACCAAATACAGCCCGTCCT
TCCAAGGCCAGGTCACCATCTCAGCCGACAAGTCCATCAACACCGCCTACC

TGCAGTGGAGCAGCCTGAAGGCCTCGGACACCGCCATGTATTACTGTGCGA
GGACC

PT8

TNGGGTTTTCCTTGTTGCTATTTTAAAGGTGTCCAGTGTGAGGTGCAGTTGG
TGGAGTCTGGGGGAGGCTCAGTTCAGCCTGGGGGGTCCCTGAGACTCTCCT
GTGGAGCCTCTGGATTCATCTTCAGGAACTACTGGATGCACTGGGTCCGCCA
AGTTCCAGGGAAGGGGCTGGTGTGGGTCTCACGCATTGATAGTGACGGTAT
TGGGACAACTCACGCGGACTCCGTGAAGGGCCGATTCATCACCTCCAGAGA
CAACGCCAAGAACACGCTGTATCTGCAAATGAACAGTCTGAGAGTCGAGGA
TATGGCTGTGTATTATTGTGCAAGAGGACACTACTACGATATGGATGTCTGG
GGCCAAGGGACCACGGTCACCGTCTCCTCAGGTA

PT9

CTCTTCGTGGTGGCAGCAGCTACAGGTGTCCAGTCCCAGGTGCAGCTGGTGC
AGTCTGGGGCTGAGGTGAAGAAGCCTGGGTCCTCGGTGAAGGTCTCCTGCA
AGGCTTCTGGAGGCACCTTCAGCAGCTATGCTATCAGCTGGGTGCGACAGG
CCCCTGGACAAGGGCTTGAGTGGATGGGAGGGATCATCCCTATCTTTGGTA
CAGCAAACTACGCACAGAAGTTCCAGGGCAGAGTCACGATTACCGCGGACG
AATCCACGAGCACAGCCTACATGGAGCTGAGCAGCCTGAGATCTGAGGACA
CGGCCGTGTATTACTGTGCGAGACCTGCGGATAGCAGCTCGTCCATATCCTA
CTACTACGGTATGGACGTCTGGGGCCAAGGGACCACGGTCACCGTCTCCTC
AGGTA

PT10

CCTTGTCGCTACTATAAAAGGTGTCCAGTGTCACGTGCAAGTGGTGGAGTTT

GGGGGAGGGTTGGTCAAGTCTGGAGGGTCCCTGACACTCTCCTGTGCAGCC

TCGGGATTCAGCTTCAGTGACTACCAGATGACTTGGGTCCGCCAGCCTCCAG

GGAAGGGGCTGGAGTGGATTTCAGACATTAGCAGTAGTAGTACTTACACNG

AAATACGCAGAATCTGTGAAGGGCCGATTCACCATCTCCAGAGACAACGCC

AAGAATTTATTGTATCTGCAAATGAACAACGTGAGAGCCGATGACACGGCT

GTGTATTATTGTGCGAGAGGCAAATATTATGGTAGTAGTGGCACTTATTACT

TTGACAACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGTA

PT11

CCTCCTCCTCGGTGGCAGCTCCCAGATGGGTCCTGGTCCAGGTGCAGCTGCA

GGAGTCGGGCGTACGACTGGTGCAATGATGGTTGATATTGCATGTGGACCC

GGCAGTCCCCAGGGGAGGGACTGGAGTGCATTGAATCTATCTAATATTCTG

GGAGCATCAACTACAATCCCTCCCTCAGGAGTCGAGTCACCGTTTCACTAAG

GACGTCCAAGAATTAGTTCTCCCTGAAGTTGAGGTCTGTGACCGCTGCGGAC

ACCGTCGTATATTCGTGTGCGAGAGAGAGGTTGGGGGTTGTCGTAGTGGTG

GTAGCTGCTACGGACGAACTCACAGTACTTCTACTACATGGACGTCTGGGA

CTAAGGGACCTCGGTCACCGTCTCCTCAGGTA

PT12

TATCTTCTTGGTGGCAGCAGCCACAGGAGCCCACTCCCAGGTGCAGCTGGT

GCAGTCTGGGGCTGAGGTGAAGAAGCCTGGGGCCTCAGTGAAGGTCTCCTG

CAAGGCTTCTGGATACACCTTCACCGGCTACTATATGCACTGGGTGCGACAG

GCCCCTGGACAAGGGCTTGAGTGGATGGGATGGATCAACCCTAACAGTGGT

GGCACAAACTATGCACAGAAGTTTCAGGGCTGGGTCACCATGACCAGGGAC

ACGTCCATCAGCACAGCCTACATGGAGCTGAGCAGGCTGAGATCTGACGAC

ACGGCCGTGTATTACTGTGCGAGAGGACCTATTTGTAGTGGTGGTAGCTGCT

ACTCGGACCCACCCTGGTACTACTACTACGGTATTGGACGTCTGGGGCCAA

GGGACCACGGTCACCGTCTCCTCAGGTA

PT13

TTCTCTTTGTGGTGGCAGCAGCTACAGGTGTCCAGTCCCAGGTCCAGCTGGT

GCAGTCTGGGGCTGAGGTGAAGAAGCCTGGGTCCTCGGTGAAGGTCTCCTG

CAAGGCTTCTGGAGGCACCTTCAGCAGCTATGCTATCAGCTGGGTGCGACA

GGCCCCTGGACAAGGGCTTGAGTGGATGGGAAGGATCATCCCTATCCTTGG

TATAGCAAACTACGCACAGAAGTTCCAGGGCAGAGTCACGATTACCGCGGA

CAAATCCACGAGCACAGCCTACATGGAGCTGAGCAGCCTGAGATCTGAGGA

CACGGCCGTGTATTACTGTGCGAGCTCTTACGATTTTTGGAGTGGTTATCCC

CCTTTCCTTGACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCNTAGGTA

PT14

CCTCTCCTGGCTGTTCTCCAGGAGTCTGTGCCGAGGTGAGCTGGTGCAGTCT

GGAGCAGAGGTGAAAAAGCCCGGGGAGTCTCTGAAGATCTCCTGTAAGGGT

TCTGGATACANCTTTACCACCTACTGGATCGGCTGGGTGCGCCAGATGCCCG

GGAAAGGCCTGGAGTGGATGGGGATCATCTATCCTGGTGACTCTGATNCCA

GATACAGCCCGTCCTTCCAAGGCCAGGTCACCATCTCAGCCGACAAGTCCA

TCANNACCGCCTACCTGCAGTGGAGCAGCCTGAAGGCCTCNGACACCGNCA
TGTATTACTGTGCGAGAC

PT15

GGTTTCTCCTCGTTGCTCTTTTAAGAGGTGTCCAGTGTCACATGCAGCTGGT
GGAGTCTGGGGGAGGCGTGGTCCAGCCTGGGGGATCCCTGACACTCTCCTG
TGCAGCCTCTGGATTCACCTTCAGTGGCTATGGCATGCACTGGGTCCGCCAG
GCTCCAGGCAAGGGGTTAGAGTGGGTGGCAGTTATTTCATTTGACGGAAGT
AGTGAATACTATGAAAACTCCGTGAGGGGCCGATTCACCGTCTCCAGAGAC
AATTCCAAGAGCACACTGTATCTGCAAATGGACAGCCTGAGACCTGAGGAC
ACGGCTGTATTTTACTGTGCGAAAGGCCAAAGAGAATTTGGGGGAGCCCCT
GACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGTA

PT16

CGCCTCCTCCTGGCTGTTCTCCAAGGAGTCTGTGCCGANGTGCAGCTGGTGC
AGTCTGGAGCAGAGGTGAAAAAGCCCGGGGAGTCTCTGAAGATCTCCTGTN
AGGNTTCTGGATACANNTTTACCACCTACTGGATCNNCTGGGTGCGCCAGA
TGCCCGGGAAAGGCCTGGAGTGGATGGGGANCATNNATCCTGGTGACTCTG
ATACCANATACAGCCCGTCCTTCCAAGGCCANGTCACCATCTCAGCCGACA
AGTCCATCANCACCGCCTACCTGCAGTGGAGCAGCCTGAAGGCCTCGGACA
CCGCCATGTATTACTGTGCGAGNCNNNGTTATNNNGATGNGGNTTATCNAN
ACTACTGCGGCCANGGAACCCAGGTCACCATCTCCTTAAGTACCT

PT17

TGGTTTTCCTTGTTGTTATTATAAAAGGTGTCCAGTGTCAGGTGCAGCTGGT

GGAGTCTGGGGGAGGCTTGGTCAAGCCTGGAGGGTCCCTGAGACTCTCCTG

TGCAGCCTCTGGATTCACCTTCAGTGACTACTACATGAACTGGATCCGCCAG

GCTCCAGGGAAGGGGCTGGAGTGGGTTTCATACATTAGTAGTAGTGGTAGT

ACCATATACTACGCAGACTCTGTGAAGGGCCGATTCACCATCTCCAGGGAC

AACGCCAAGAACTCACTGTATCTACAAATGAACAGCCTGAGAGCCGAGGAC

ACGGCCGTGTATTACTGTGCGAGAGCTGTTTCTCGTAGTACCAACTGCTATG

ACGACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGG

PT18

TCTCCCTGGTGGCAGCTCCCAGATGGGTCGTGTCCCAGGTGCAGCTACAACA

GTGGGGCGCAGGACTATTGAAGTCTTCGGAGACCCTGTCCCTCACCTGCGCT

GTCTATGGTGGGACCTTCAGTGGTTACTACTGGGGCTGGATCCGCCAGCCCC

CAGGGAAAGGGCTGGAGTGGATTGGGGAAATCAATCATGTTGGAAGGAAC

AACCACAGCCCGTCCCTCAAGAGCCGAGTCACCATGTCAGTAGACACGTCC

AAGAACCAATTCTCCCTGAAGCTGAGCTCTGTGACCGCCGCGGACACGGCT

CTCTATTACTGTGCGAGGTATAAGGGATATAGTGACTACGATTATTACTTTG

ACTACTGGAGCCAGGGAACCCTGGTCACCGTCTCCNTCAGGTA

PT19

CTCCTGCTGGTGGCGGCTCCCAGATGGGTCCTGTCCCAGCCTGCAACTGCAG

GAGTCGGGCCCAGGGCTGGTGCAGCCTTCGGAGACCCTGTCCCTCACCTGC

GCTGTCTCTGGTGCCTCCATCAGCACTAATAATTACTACTGGGGCTGGGTCC

GCCAGCCCCCAGGGAAGGGGCTGGAGTGGATTGGGTGTATCCATCATAGTG

GGAGCACCTACTATAGTCCGTCCCTCAGGAGTCGAGTCATCATGTCCGTAGA

CACGTCCAAGAACCAGTTCTCTCTGAATCTGAACTCTGTGACCGCCGCAGAC

ACGGCTGTGTATTACTGTGCGAAACAGGGTGGCGCCTCTTCCGGGGACTACT

GGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGTAC

PT20

TGGGTTCTCCTCGTTGCTCTTTTAAGAGGTGTCCAGTGTCAGGTGCAGCTGG

TGGAGTCTGGGGGAGCCGTGGTCCAGCCTGGGAGGTCCCTGAGACTCTCCT

GTGCAGCGTCTGGATTCACCTTCACTAACTATGGCATGCACTGGGTCCGCCA

GGCTCCAGGCAAGGGGCTGGAGTGGGTGGCACTTATATGGTTTGATGGAAG

TACTAAATACTATGCAGACTCCGTGAAGGGCCGATTTACCATCTCCAGAGA

CAATTCCAAGAACACGCTGTTTCTGCAAATGAAAAGCCTGAGAGCCGAGGA

CACGGCTGTGTATTACTGTGCGAGAGATACCTTAGGTTACCGTGATAGTCGT

GGGTCTAACGGAGGTGTCTACTGGGGCCAGGGAACCCTGGTCACCGTCTCC

TCAGGTA

PT21

CTCTCTTGGTGGCAGCAGCCACAGGAGCCCACTCCCAAGGTGCACGCTGGT

GCAGTCTGGGGCTGAGGTGAAGAAGNCCTGGGGCCTCAGTGAAGGTCTCCT

GCAAGGCTTCTGGATACACCTTCACCGACTACTNTATGCACTGGGTGCGACA

GGCCCCTGGACAAGGGCTTGAGTGGATGGGACGGATCAACCCTGACNGTGG

TGACACAAACTATGCACNGAAGTTTCAGGACAGGGTCACCATGACCAGGGA

CACGTCCATCAGNACAGCCTACATGGANCTGAGCAGGCTGACATCTGACGA

CACGGCCGTGTATTANTGTGCGAGAGCCCACTCTTATTATAGTAGGNAC

PT22

CCTCCTCCTGGTGGCAGCTCCCAGATGGGTCGTGTCTCAGGTGCAGCTGCAG

GAGTCGGGCCCAGGACTGGTGCAGCCTTCGGGGACCCTGTCTCTCACCTGC

GCTGTCTCTGATGGCTCCATCAGCAGTAATAACTGGTGGTGGAGTTGGGTCC

GCCAGCCCCCAGGGAAGGGGCTGGAGTGGATTGGGGACATCTGTCATAGTG

GGTACACCAACTACAACCCGTCCCTCATGAGTCGAGTCACCATGTCAGTAG

ACAAGTCCAAGAATCAGTTCTCCCTGAAGGTGACCTCTGTGAGCGCCGCGG

ACACGGCCGTATATTACTGTGCGAGAGCCCTCGCATACTGTGGTGGGGACT

GCTATCCCACATACTACTACTACCACGGTATGGACGTCTGGGGCCAGGGGA

CCACGGTCACCGTCTNCTTCAGGTA

PT23

GGGGTTACCTCGTTGCTCTTTTAAGAGGTGTCCAGCGTCACGTGTAGTTGGT

GGAGTCTGGGGGAGGTGTGGTCCAGCCTGGGGGGTCCCTGAGAGTCTCCTG

TGCAGCCTCGGCATTCACCCTCAGAAGCTGTGGGGTCCGCCAGGCTCCAGG

CAAGGGACTGGAGTGGGTGGCATGTCTCATTTAATGGAAGTAACAAATATT

GATGCGAAAGATTCATCTGTCGGTGACAACCCCAAATTAGATTCATCTGTCG

GTGACAACCCCAAATTAGGACGAAGTCCTACTTTGCCATGGACGTCTGGGG

CCAAGGGACCACGGTCACCGTCTCCTTCAGGTAA

PT24

TATCTTTTTGGTGGCAGCAGCCACAGGTGCCCACTCCCAGGTCCAGCTTGTG

CAGTCTGGGGCTGAGGTGAAGAAGCCTGGGGCCTCAGTGAAGGTTTCCTGC

AAGGCTTCTGGATACACCTTCACTAGCTATGCTATGCATTGGGTGCGCCAGG

CCCCCGGACAAAGGCTTGAGTGGATGGGATGGATCAACGCTGGCAATGGTA

ACACAAAATATTCACAGAAGTTCCAGGGCAGAGTCACCATTACCAGGGACA

CATCCGCGAGCACAGCCTACATGGAGCTGAGCAGCCTGAGATCTGAAGACA

CGGCTGTGTATTACTGTGCGAGAGAACAGTGGCTGGTACGGGTCTACTTTGA

CTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGTA

PT25

CCTCCTCCTGGTGGCAGCTCCCAGATGGGTCCTGTCCCAGGTGCAGCTACAG

CAGTGGGGCGCAGGACTGTTGAGGCCCTCGGAGACCCTGTCCCTCACCTGC

GCTGTGTATGGTGGGTCCTTCAGTGGTTTCTATTGGAGCTGGATCCGCCAGC

CCCCCGGAAAGGGGCTGGAGTGGATTGGGGAAATCAATCATAGTGGAAGC

GCCAACTACAACCCGTCCCTCAAGAGTCGAGTCACCATATCATTAGACACG

TCGAAGAGTCAGTTCTCCCTGAGGCTGAGCTCTGTGACCGCCGCGGACACG

GCTGTGTATTACTGTGCGAGAGGCGAGGGTTCGGGGAGTTATTATAGTTACT

TTGACTACTGGGGCCGGGGAACTCTGGTCACCGTCTCCTTAAGGT

PT26

TGGNTTTCCTTTTGGCTATTTTAAAAGGTGTCCAGTGTGAAGTGCAGCTGGT

GGAGTCTGGGGGAGGCTTGGTACAGCCTGGCAGGTCCCTGAGACTCTCCTG

TGCAGCCTCTGGATTCACCTTTCATGATTATGGCATGCACTGGGTCCGGCAA

GCCCCAGGGAGGGGCCTGGAGTGGGTCGCAGGTATTACTTGGAATAGTGGT

TACATAGACTATGCGGACTCTGTGAAGGGCCGATTCACCATCTCCAGAGAC

AACGCCAAGAACTCCCTGTATCTGCAAATGAACAGTCTGAGCGCTGAGGAC

ACGGCCTTGTATTACTGTGCAAAAGTTGGAGGCAGTACCTGGTCAAATATTG
ACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGTA

PT27

TGGTTTTCCTTGTTGCTATTTTAAAAGGTGTCCAGTGTGAGGTGAGCTGGTG
GATCTGGGGGAGGCTTGGTACAGCCTGGGGGGTCCCTGAGACTCTCCTGTG
CAGCCTCTGGATTCACCTTCAGCAGCTATGCCATGAACTGGGTCCGCCAGGC
TCCAGGGAAGGGGCTGGAGTGGGTTTCATACATTAGTAGTAGTGGTGGAAC
CATATATTACGCAGACTCTGTGAAGGGCCGATTCACCATCTCCAGAGACtAT
GCCAATAACTCACTGTTTCTGCAAATGAACAGCCTGAGAGACGAGGACACG
GCTGTATATTACTGTGCGAGAGGCCCAGGCTACGATTACGTTTGGGGGAATT
ATCGCCCGACCCTTTACTTTGACTACTGGGGCCAGGGAACCCTGGTCACCGT
CTCTCAGGTA

PT28

TGGTTTCCTCTTGTTGTATTTTAAAAGGTGTCCAGTGTGAGGTGCAGCTGGT
GGAGTCTGGGGGAGGCTTGGTACAGCCTGGGGGGTCCCTGAGACTCTCCTG
TGCAGCCTCTGGATTCACCTTCAGTAACTATTGCATGAACTGGGTCCGCCAG
GCTCCAGGGAAGGGGCTGGAGTGGGTCTCACGTATTAGTAGTGATGGTAGT
AATAAATACTATGCATACTCTGTGAAGGGCCGATTCACCATCTCCAGAGAC
AATTCCAAGAACACGCTGTATCTGCAAATGAACAGGCTGAGAGCCGAGGAC
ACGGCTGTGTATTACTGTGCGAAAGAT

PT29

GGGTTTTCCTCGTTGCTCTTTTAAGAGGTGTCCAGTGTCAGGTGCAGCTGGT

GGAGTCTGGGGGAGGCGTGGTCCAGCCTGGGAGGTCCCTGAGACTCTCCTG

TGCAGCGTCTGGATTCACCTTCAGTAGTTATGGCATGTACTGGGTCCGCCAG

GCTCCAGGCAAGGGGCTGGAGTGGGTGGCAGTTATATGGTGTAATGGAAGT

AAAAAATACTATGCAGACTCAGTGAAGGGCCGATTCACCATCTCCAGAGAC

AATCCCAAGAATATGGTATGTCTGCAAATGAACAGTCTGAGAGTCGAGGAC

ACGGGTGTGTATTAATGTGCGAGAGATCCTTTGTGTTCGGTTGGTGGTAATG

GCTACAGGCTACTACCACGGTATGGACGTCTGGGGCCAAGGGACCCGGTCA

CCGTTTCTTCAGGTA

PT30

TATCTTCTTGGTGGCAGCAGCCACAGGAGCCCACTCCCAGGTGCAGCTGGT

GCAGTCTGGGGCTGAGGTGAAGAAGCCTGGGGCCTCAGTGAAGGTCTCCTG

CAAGGCTTCTGGATACACCTTCACCGGCTACTATATGCACTGGGTGCGACAG

GCCCCTGGACAAGGGCTTGAGTGGATGGGATGGATCAACCCTAACAGTGGT

GGCACAAACTATGCACAGAAGTTTCAGGGCAGGGTCACCATGACCAGGGAC

ACGTCCATCAGCACAGCCTACATGGAGCTGAGCAGGCTGAGATCTGACGAC

ACGGCCGTGTATTACTGTGCGAGAGATCAATGGTTCGGGATTAACTACTTTG

ACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGTC

Appendix D: Summary of patient information.

Table D.1. Summary of patient information.

					CDR1+CDR2			FR1+FR2+FR3		
Sample	**MS**	**V_H family**	**V_H gene**	**% ID**	**N2**	**R2/S2**	**P2**	**N1**	**R1/S1**	**P1**
PT1	M	VH4	V4-34	92.02%	10	7/3	0.0698	13	7/6	0.0056
PT2	M	VH5	V5-51	95.83%	3	2/1	0.5143	9	7/2	0.4806
PT3	UM	VH3	V3-74	99.40%	1	1/0	0.1833	1	1/0	0.4291
PT4	UM	VH3	V3-33	100%	0	0/0	0	0	0/0	1.0
PT5	M	VH3	V3-23	94.60%	8	5/3	0.0902	8	3/5	0.001
PT6	M	VH5	V5-51	91.90%	5	5/0	0.0257	7	5/2	0.1538
PT7	UM	VH5	V5-51	99%	1	1/0	0.1767	1	1/0	0.4082
PT8	M	VH3	V3-74	92.52%	9	6/3	0.1479	13	8/5	0.0279
PT9	UM	VH1	V1-69	99.67%	0	0/0	0.5876	1	0/1	0.2101
PT10	M	VH3	V3-11	91.10%	10	8/2	0.1022	18	12/6	0.0581
PT11		V4	-	-	-	-	-	-	-	-
PT12	UM	VH1	VH1-2	100%	0	0/0	0	0	0/0	1.0
PT13	UM	VH1	VH1-69	98.70%	3	3/0	0.00980	1	0/1	0.01536
PT14	M	VH5	VH5-51	97.62%	2	2/0	0.17380	4	4/0	0.32905
PT15	M	VH3	VH3-30	92.80%	9	7/2	0.04432	12	7/5	0.01449
PT16	M	VH5	VH5-51	95.30%	5	4/1	0.03047	6	4/2	0.04111
PT17	UM	VH3	VH3-11	99.32%	1	1/0	0.18247	1	0/1	0.09079
PT18	M	VH4	VH4-34	94.40%	7	6/1	0.03408	9	4/5	0.00609
PT19	M	VH4	VH4-39	93.70%	10	8/2	0.02146	11	8/3	0.000
PT20	M	VH3	VH3-33	96.94%	4	4/0	0.03843	5	4/1	0.22674
PT21	M	VH1	VH1-2	95.92%	8	7/1	0.00130	4	3/1	0.01329
PT22	M	VH4	VH4-28	91.31%	7	7/0	0.02418	17	10/7	0.02874
PT23		VH1	-	-	-	-	-	-	-	-
PT24	M	VH1	VH1-3	97.96%	2	2/0	0.19182	4	1/3	0.02691
PT25	M	VH4	VH4-34	95.50%	3	2/1	0.55921	10	4/6	0.03206
PT26	M	VH3	VH3-9	96.40%	5	4/1	0.02584	5	4/1	0.09563
PT27	M	VH3	VH3-48	96.02%	5	3/2	0.07581	4	3/1	0.05312
PT28	M	VH3	VH3-11	95.92%	8	7/1	0.001	4	4/0	0.05378
PT29	M	VH3	VH3-33	95.24%	6	4/2	0.166	8	6/2	0.13545
PT30	UM	VH1	VH1-2	100%	0	0/0	0	0	0/0	1.0

CDR = complementarity-determining Region; FR = framework region;
R = number of detected R mutations; S = number of detected S mutations;
MS= mutational status.

Appendix E: Quality of the isolated small RNA.

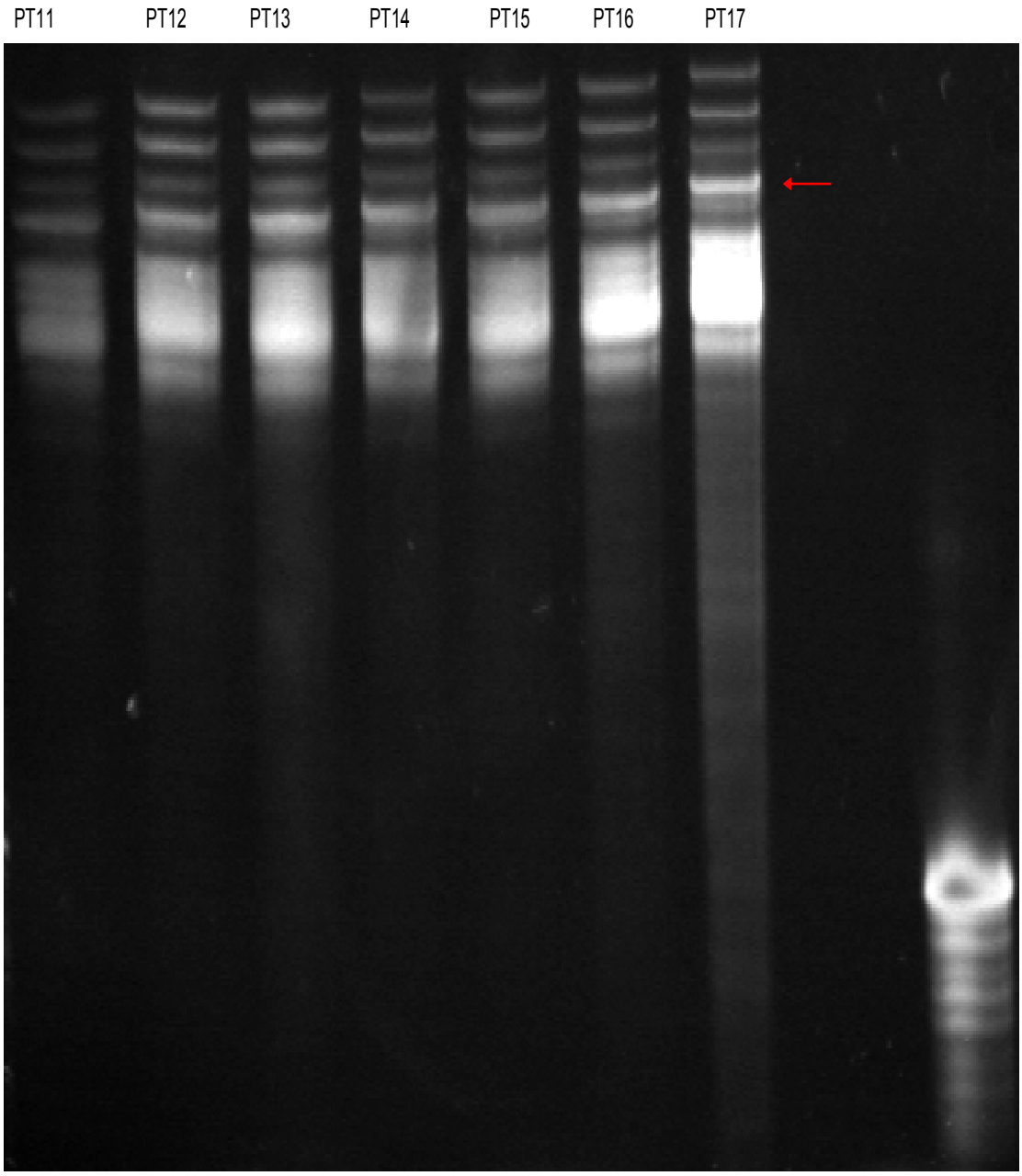

Figure E.1. Quality of the small isolated RNA from PT11-PT17. miRNA concentration of 0.25µg miRNA samples of PT11-PT17 and 0.25µg mir-16 probe analyzed on a 15 % polyacrylamide gel amd stained with cybergold.

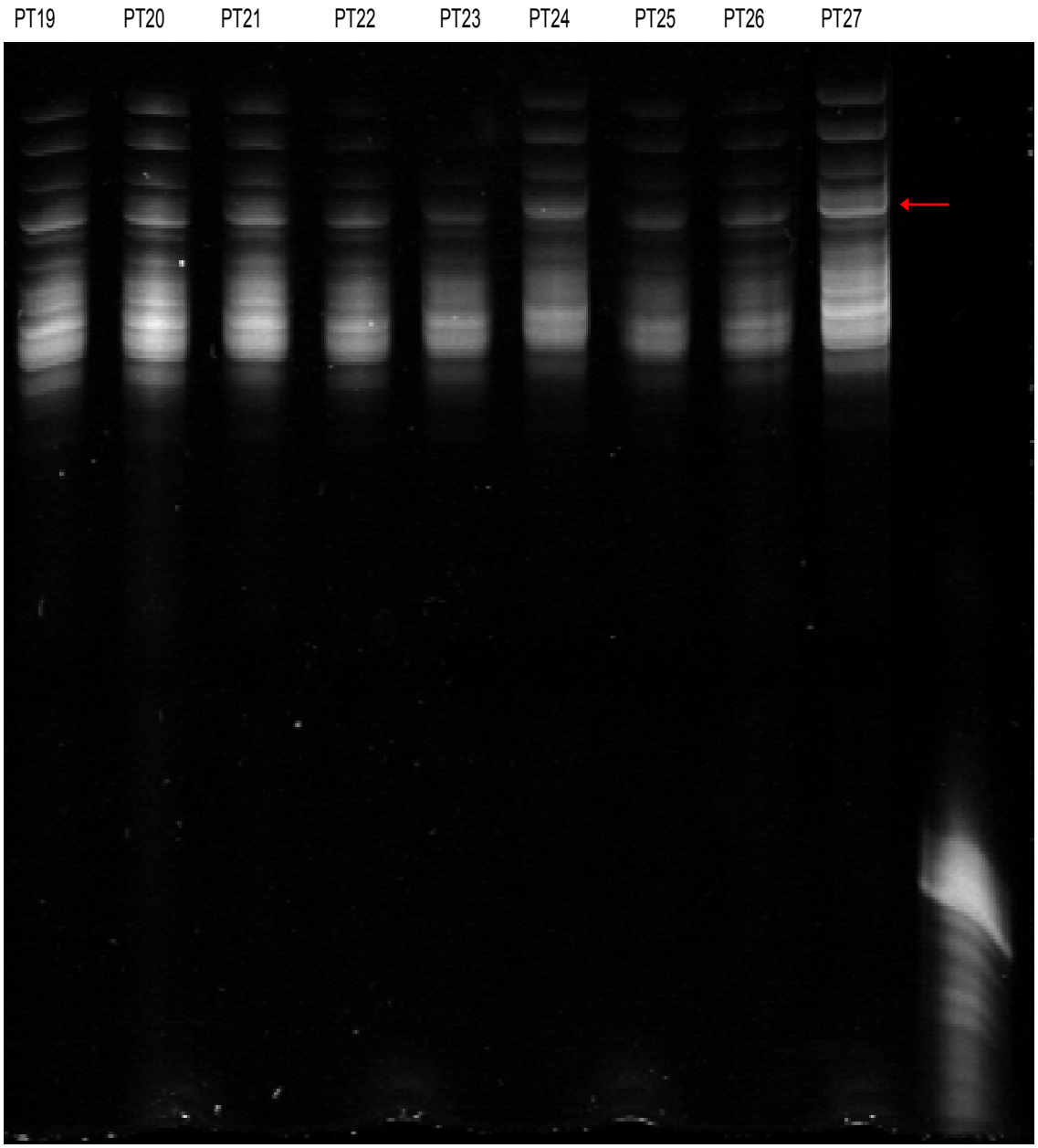

Figure E.2. Quality of the small isolated RNA from PT19-PT27. miRNA concentration of 0.25µg miRNA samples of PT19-PT27 and 0.25µg mir-16 probe analyzed on a 15 % polyacrylamide gel amd stained with cybergold.

.

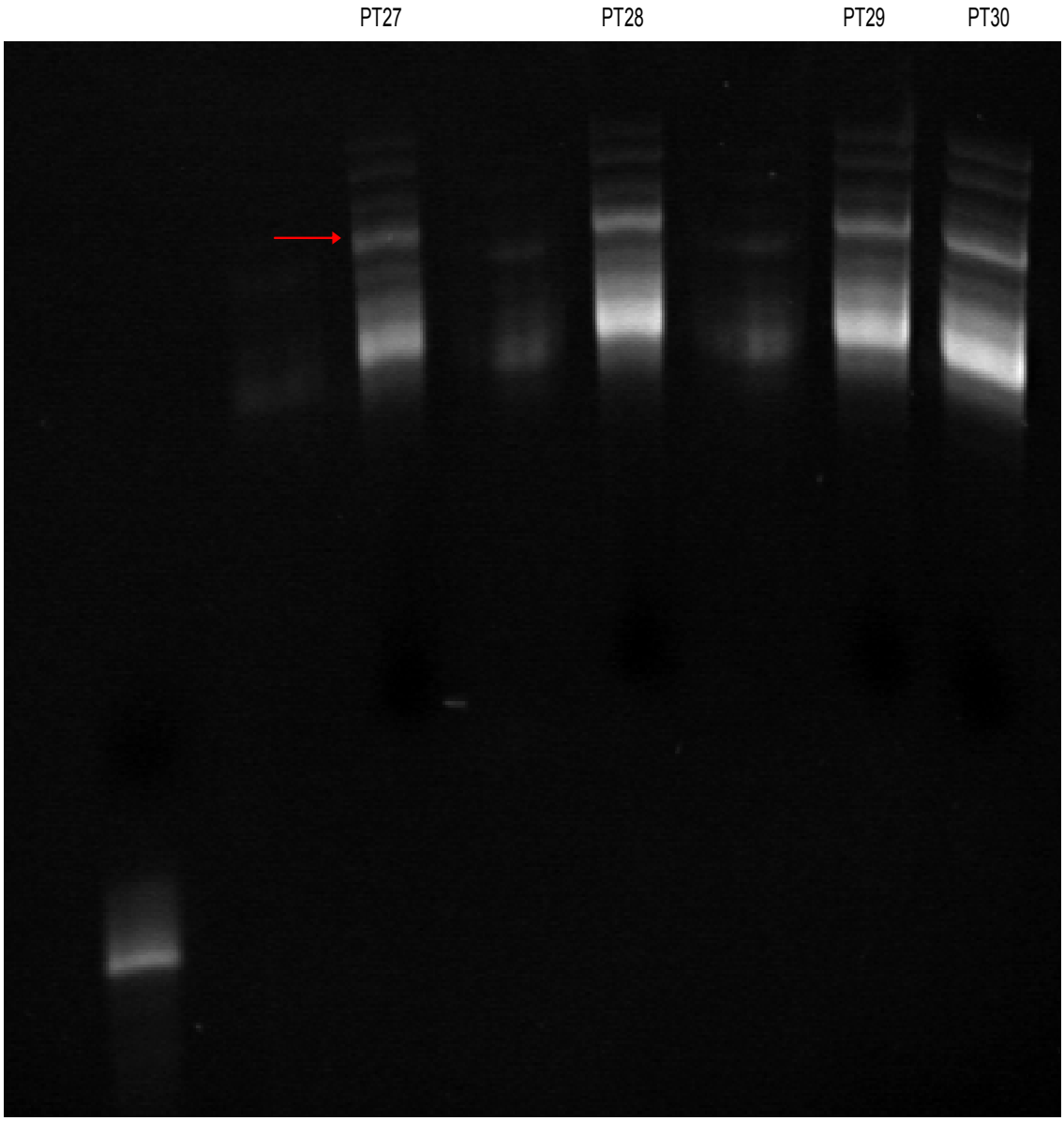

Figure E.3. Quality of the small isolated RNA from PT27-PT30. miRNA concentration of 0.25µg miRNA samples of PT27-PT30 and 0.25µg mir-16 probe analyzed on a 15 % polyacrylamide gel amd stained with cybergold.

.

Appendix F: Seventy-two selected miRNAs.

Table F.1. Seventy-two miRNAs that are located near B-CLL chromosome aberrations.

Aberration	miRNA	Aberration	miRNA
d6p21	hsa-miR-219	**+3q13.33**	hsa-mir-198
d6q12.2	hsa-miR-206	**+3q25.33**	hsa-mir-15b
d6q13	hsa-miR-30c	**+3q25.33**	hsa-mir-16-2
d6q13	hsa-miR-30a	**+3q28**	hsa-mir-28
	hsa-miR-30a*		
		+ 8p23.1	hsa-miR-124a-1
d11q13	hsa-miR-192		
	hsa-miR-194-2	**+ 8p12.3**	hsa-mir-124a-2
	hsa-miR-139		
		+ 8p21.3	hsa-miR-320
d11q23	hsa-miR-34b		
		+ 8q24.2	hsa-miR-30b
d11q23	hsa-miR-34c		
		+ 8q24.22	hsa-miR-30d
d11q24	hsa-miR-125b		
		+12q13.13	hsa-miR-196-2
d11q24	hsa-let-7a		
		+12q13.31	hsa-miR-200c
d11q24	hsa-miR-100	**+12q13.31**	hsa-miR-141
d13q14	hsa-miR-16-1	**+12q14.1**	hsa-miR-26a-2
d13q14	hsa-mir-15a	**+12q14.2**	hsa-let-7i
d17p13.3	hsa-miR-22	**+12q23.1**	hsa-miR-135-2
d17p13.3	hsa-miR-132		
		+19q13.2	hsa-miR-199a
d17p13.3	hsa-miR-212		
			hsa-miR-199a*
d17p13.1	hsa-miR-195		
		+19q13.2	hsa-miR-24-2
d17q11.2	hsa-miR-144		
		+19q13.2	hsa-miR-27a
d17q11.2	hsa-miR-193		
		+19q13.2	hsa-miR-23a
d17q21.32	hsa-miR-152		

d17q21.32	hsa-miR-10a
d17q21.32	hsa-miR-196-1
d17q23.2	hsa-miR-142-5p hsa-miR-142-3p
d17q23.2	hsa-miR-301
d17q23.2	hsa-miR-21
d18q 11.2	hsa-miR-133a-1
d18q 11.2	hsa-miR-1-2
+3p22.3	hsa-miR-128b
+3p22.3	hsa-miR-26a-1
+3p21.33	hsa-mir-138-1
+3p21.31	hsa-mir-191
+3p21.2	hsa-let-7g
+3p21.2	hsa-mir-135-1
+19q13.3	hsa-miR-7-3
+19q13.3	hsa-miR-181c
+19q13.33	hsa-miR-150
+19q13.4	let-7e
+19q13.4	miR-099b
+19q13.4	miR-125a
+19q13.4	miR-hes1 miR-hes2 miR-hes3
t 14q32.31	hsa-miR-127 hsa-miR-136 hsa-miR-299 hsa-miR-134 hsa-miR-154
t 14q32.33	hsa-miR-203
t 18q21	hsa-miR-122a
	72 total

Appendix G: Thirty selected miRNAs.

Table G.1. Selected 30 miRNAs: their mature sequences and DNA oligonucleotide templates

miRNA	miRNA Sequence (Mature)	DNA oligonucleotide for T7 miRNA Probe (+4Ts and 0T)
hsa-miR-194-2	15uguaacagcaacuccaugugga36	Tgtaacagcaactccatgtgga----**cctgtctc**
hsa-miR-34**b**	14aggcagugucauuagcugauug35	aggcagtgtcattagctgattgtttt**cctgtctc**
hsa-miR-125b	15ucccugagacccuaacuuguga36	tccctgagaccctaacttgtga----**cctgtctc**
hsa-let-7a	5ugagguaguagguuguauaguu26	tgaggtagtaggttgtatagttttt**cctgtctc**
hsa-miR-16-1	14uagcagcacguaaauauuggcg35	tagcagcacgtaaatattggcg----**cctgtctc**
hsa-mir-15a	14uagcagcacauaaugguuugug35	tagcagcacataatggtttgtgttt**cctgtctc**
hsa-miR-22	53aagcugccaguugaagaacugu74	Aagctgccagttgaagaactgt----**cctgtctc**
hsa-miR-132	59uaacagucuacagccauggucg80	taacagtctacagccatggtcgttt**cctgtctc**
hsa-miR-212	71uaacagucuccagucacggcc91	taacagtctccagtcacggcc----**cctgtctc**
hsa-miR-195	15uagcagcacagaaauauuggc35	tagcagcacagaaatattggcttt**cctgtctc**
hsa-miR-133a	54uuggucccccuucaaccagcugu75	Ttggtccccttcaaccagctgt----**cctgtctc**
hsa-miR-124a	52uuaaggcacgcggugaaugcca73	ttaaggcacgcggtgaatgccattt**cctgtctc**
hsa-miR-320	48aaaagcuggguugagagggcga70	Aaaagctgggttgagagggcgaa----**cctgtctc**
hsa-miR-30b	17uguaaacauccuacacucagc37	tgtaaacatcctacactcagctttt**cctgtctc**
hsa-miR-30d	6uguaaacauccccgacuggaag27	Tgtaaacatccccgactggaag----**cctgtctc**
hsa-miR-196-2	25uagguaguuucauguuguuggg46	taggtagtttcatgttgttgggtttt**cctgtctc**
hsa-miR-200c	45aauacugccggguaaugaugga66	Aatactgccgggtaatgatgga----**cctgtctc**
hsa-miR-141	60aacacugucugguaaagaugg80	aacactgtctggtaaagatggtttt**cctgtctc**
hsa-miR-26a-2	14uucaaguaauccaggauaggcu35	Ttcaagtaatccaggataggct----**cctgtctc**
hsa-let-7i	6ugagguaguaguuugugcu24	tgaggtagtagtttgtgttttt**cctgtctc**
hsa-miR-135-2	23uauggcuuuuuauuccuaugug45	Tatggcttttattcctatgtga----**cctgtctc**
hsa-miR-122a	15uggagugugacaaugguguuug37	tggagtgtgacaatggtgtttgttttt**cctgtctc**
hsa-miR-199a	6cccaguguucagacuaccuguu27	Cccagtgttcagactacctgtt----**cctgtctc**
hsa-miR-24-2	50uggcucaguucagcaggaacag71	tggctcagttcagcaggaacagtttt**cctgtctc**
hsa-miR-27a	51uucacaguggcuaaguuccgcc72	Ttcacagtggctaagttccgcc----**cctgtctc**
hsa-miR-23a	45aucacauugccagggauuucc65	atcacattgccagggatttcctttt**cctgtctc**
hsa-miR-7-3	31uggaagacuagugauuuuguu51	Tggaagactagtgattttgtt----**cctgtctc**
hsa-miR-181c	27aacauucaaccugucggugagu48	aacattcaacctgtcggtgagttttt**cctgtctc**
let-7e	8ugagguaggagguuguauagu28	Tgaggtaggaggttgtatagt----**cctgtctc**
miR-125a	15ucccugagacccuuuaaccugu37	tccctgagacccttt aacctgtgtttt**cctgtctc**

Appendix H: Eight miRNAs chosen for expression analysis.

Table H.1. Eight chosen miRNAs andtheir target mRNAs.

miRNA	Targets (human)	Function
Deletion 13:		
mir16-1	ENSG00000158796	Death effector Domain-containing protein (FLDED-1), (KE05).
Deletion 11:		
mir-34b	ENSG00000148400	Neurogenic locus notch homolog protein 1recursor (NOTCH 1), (Translocation associated notch protein TAN-1)
mir-125b	ENSG00000126929	HIV-1 inducer of short transcripts binding protein; Lymphoma related factor
	ENSG00000128342	Leukemia inhibitory factor recursor (LIF), (Differentiation stimulating factor), (D factor), (Melanoma derived LPL inhibitor), (MLPLI).
	ENSG00000104081	BCL-2 modifying factor.
Trisomy19:		
mir-181C	ENSG00000169635	Hypermethylated in cancer 2 protein (HIC-2), (HIC-3), (HIC1 related gene on chromosome 22)
mir-125a	ENSG00000116017	Dead ringer like-1 protein (B-cell regulator of IGH transcription), (Bright).
	ENSG00000126929	HIV-1 Inducer of short transcripts binding protein; lymphoma related factor.

	ENSG00000128342	Leukemia inhibitory factor precursor (LIF), (Differentiation stimulating factor), (D factor), (Melanoma derived LPL inhibitor)
Trisomy12:		
let-7i	ENSG00000169635	Hypermethylated in cancer 2 protein (HIC-2), (HIC3),(HIC1-related gene on Chromosome 22)
mir196-2	ENSG00000148200	Orphan nuclear receptor NR6A1 (Germ cell nuclear factor), (GCNF),(retinoid receptor related testis specific receptor).
mir26a-2	ENSG00000170365	Mothers against decapentaplegic homolog (SMAD 1), (Mothers against DPP homolog 1), (MAD related protein 1), (Transforming growth factor beta signaling protein 1),(BSP-1), (HSMAD1),(JV4-1).

Appendix I: Output of SAS and the analysis process.

The following is result of the output of SAS and the analysis process:

mir-34b

(1) Intensity or expression

Table I.1. The expressions of mir-34b in the two groups (UM, M)

Rep	M	UM
1	148019	
2	107395.7	
3	23281.6	
4	66177.7	
5	48349.5	
6	168044.3	
7	29177.4	
8	25258.4	
9		242040.9
10		159247.7
11		142657.1
12		121079.5

(2) The SAS input file

```
data first;
input Rep species$ intensity;
Cards;
1       M     148019
1       M     107395.7
1       M     23281.6
1       M     66177.7
1       M     48349.5
1       M     168044.3
1       M     29177.4
1       M     25258.4
2       UM    242040.9
2       UM    159247.7
2       UM    142657.1
2       UM    121079.5
;
proc ttest;
      class species;
      var   intensity;
            title 'T-test using SAS ttest';
```

Figure I.1. The SAS input file for mir-34b expression

(3) The SAS output file

T-test using SAS ttest 10:28 Friday, August 13, 2004 1

The TTEST Procedure
Statistics

Variable	patients	N	Mean	Mean	Mean	Std Dev	Std Dev	Std Dev	Std Err
intensity	M	8	28995	76963	124931	37936	57377	116777	20286
intensity	UM	4	82105	166256	250407	29958	52884	197182	26442
intensity	Diff (1-2)		-166E3	-89293	-12793	39175	56067	98393	34334

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
intensity	Pooled	**Equal**	10	-2.60	**0.0265**
intensity	Satterthwaite	Unequal	6.59	-2.68	0.0334

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
intensity	Folded F	7	3	1.18	0.9811

Figure I.2. The SAS output file for mir-34b expression

let-7i

(1) Intensity or expression

Table I.2. The expressions of let-7i in the two groups (UM, M)

Rep	M	UM
1	638895.5	
2	698378	
3	723402.6	
4	687411.5	
5	744980.2	
6	620563.6	
7	759537.4	
8	774914.3	
9		906091
10		1192292.6
11		1550496.9
12		1516398.9

(2) The SAS input file

```
data first;
input Rep species$ intensity;
Cards;
1      M      638895.5
1      M      698378
1      M      723402.6
1      M      687411.5
1      M      744980.2
1      M      620563.6
1      M      759537.4
1      M      774914.3
2      UM     906091
2      UM     1192292.6
2      UM     1550496.9
2      UM     1516398.9
;
proc ttest;
     class species;
     var   intensity;
          title 'T-test using SAS ttest';
run;
```

Figure I.3. The SAS input file for let-7i expression

(3) The SAS output file

T-test using SAS ttest 10:43 Friday, August 13, 2004 1

The TTEST Procedure

Statistics

Variable	patients	N	Mean	Mean	Mean	Std Dev	Std Dev	Std Dev	Std Err
intensity	M	8	659523	706010	752498	36765	55605	113172	19659
intensity	UM	4	808642	1.29E6	1.77E6	171838	303338	1.13E6	151669
intensity	Diff (1-2)		-821E3	-585E3	-35E4	120553	172535	302788	105656

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
intensity	Pooled	Equal	10	-5.54	0.0002
intensity	Satterthwaite	**Unequal**	3.1	-3.83	**0.0296**

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
intensity	Folded F	3	7	29.76	0.0005

Figure I.4. The SAS output file for let-7i expression

mir-26a-2

(1) Intensity or expression

Table I.3. The expressions of mir-26a in the two groups (UM, M)

Rep	M	UM
1	4271839.8	
2	5154553.9	
3	2639191	
4	3606797	
5	3216792.2	
6	3688059.3	
7	1894745.6	
8	2793187.3	
9		1421870.9
10		1568594.1
11		1803838.1
12		2179905.6

(2) The SAS input file

```
data first;
input Rep species$ intensity;
Cards;
1     M     4271839.8
1     M     5154553.9
1     M     2639191
1     M     3606797
1     M     3216792.2
1     M     3688059.3
1     M     1894745.6
1     M     2793187.3
2     UM    1421870.9
2     UM    1568594.1
2     UM    1803838.1
2     UM    2179905.6
;
proc ttest;
     class species;
     var   intensity;
          title 'T-test using SAS ttest';
run;
```

Figure I.5. The SAS input file for mir-26a expression

(3) The SAS output file

T-test using SAS ttest 10:56 Friday, August 13, 2004 1

The TTEST Procedure

Statistics

Variable	patients	N	Mean	Mean	Mean	Std Dev	Std Dev	Std Dev	Std Err
intensity	M	8	2.56E6	3.41E6	4.26E6	671027	1.01E6	2.07E6	358822
intensity	UM	4	1.22E6	1.74E6	2.27E6	187350	330720	1.23E6	165360
intensity	Diff (1-2)		479930	1.66E6	2.85E6	606650	868235	1.52E6	531683

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
intensity	Pooled	**Equal**	10	3.13	**0.0107**
intensity	Satterthwaite	Unequal	9.31	4.21	0.0021

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
intensity	Folded F	7	3	9.42	0.0924

Figure I.6. The SAS output file for mir-26a expression

mir-16-1

(1) Intensity or expression

Table I.4. The expressions of mir-16-1 in the two groups (UM, M)

Rep	M	UM
1	11304.2	
2	7078.1	
3	109980.3	
4	143141.8	
5	269461.6	
6	37943.9	
7	7116.5	
8	25721	
9		152280.7
10		256082.4
11		122779.7
12		247899

(2) The SAS input file

```
data first;
input Rep species$ intensity;
Cards;
1      M      11304.2
1      M      7078.1
1      M      109980.3
1      M      143141.8
1      M      269461.6
1      M      37943.9
1      M      7116.5
1      M      25721
2      UM     152280.7
2      UM     256082.4
2      UM     122779.7
2      UM     247899
;
proc ttest;
     class species;
     var   intensity;
          title 'T-test using SAS ttest';
```

Figure I.7. The SAS input file for mir-16-1 expression

(3) The SAS output file

T-test using SAS ttest 10:59 Friday, August 13, 2004 1

The TTEST Procedure

Statistics

Variable	patients	N	Mean	Mean	Mean	Std Dev	Std Dev	Std Dev	Std Err
intensity	M	8	-1411	76468	154348	61592	93155	189596	32935
intensity	UM	4	87742	194760	301779	38099	67255	250765	33628
intensity	Diff (1-2)		-236E3	-118E3	-667.9	60234	86206	151286	52790

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
intensity	Pooled	**Equal**	10	-2.24	**0.0489**
intensity	Satterthwaite	Unequal	8.26	-2.51	0.0353

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
intensity	Folded F	7	3	1.92	0.6377

Figure I.8. The SAS output file for mir-16-1 expression

mir-181c

(1) Intensity or expression

Table I.5. The expressions of mir-181c in the two groups (UM, M)

Rep	M	UM
1	328896.6	
2	282554.84	
3	281104	
4	362010.9	
5	371392.2	
6	433773.5	
7	243226.3	
8	407736.9	
9		279577.5
10		255559.8
11		226375.7
12		311634.8

(2) The SAS input file

```
data first;
input Rep species$ intensity;
Cards;
1      M     328896.6
1      M     282554.84
1      M     281104
1      M     362010.9
1      M     371392.2
1      M     433773.5
1      M     243226.3
1      M     407736.9
2      UM    279577.5
2      UM    255559.8
2      UM    226375.7
2      UM    311634.8
;
proc ttest;
      class species;
      var   intensity;
           title 'T-test using SAS ttest';
run;
```

Figure I.9. The SAS input file for mir-181c expression

(3) The SAS output file

T-test using SAS ttest 11:01 Friday, August 13, 2004 1

The TTEST Procedure

Statistics

Variable	patients	N	Mean	Mean	Mean	Std Dev	Std Dev	Std Dev	Std Err
intensity	M	8	283114	338837	394560	44069	66653	135656	23565
intensity	UM	4	210731	268287	325843	20491	36171	134866	18086
intensity	Diff (1-2)		-10199	70550	151299	41350	59180	103857	36240

T-Tests

Variable	Method	Variances	DF	t Value	Pr > |t|
intensity	Pooled	**Equal**	10	1.95	**0.0802**
intensity	Satterthwaite	Unequal	9.77	2.37	0.0395

Equality of Variances

Variable	Method	Num DF	Den DF	F Value	Pr > F
intensity	Folded F	7	3	3.40	0.3432

Figure I.10. The SAS output file for mir-181c expression